UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Tecnologia

Juan Fernando Galindo Jaramillo

Agent-Based Modeling and Simulation of Social Determinants of Health

Modelagem e Simulação Baseada em Agentes de Determinantes Sociais de Saúde

Limeira
2023

Juan Fernando Galindo Jaramillo

Agent-Based Modeling and Simulation of Social Determinants of Health

Modelagem e Simulação Baseada em Agentes de Determinantes Sociais de Saúde

Tese apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutor em Tecnologia, na área de Sistemas de Informação e Comunicação.

Thesis presented to the School of Technology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Technology in Computer Science, in the area of Sistemas de Informação e Comunicação.

Supervisor/Orientador: Prof. Dr. Paulo Sérgio Martins Pedro
Co-supervisor/Coorientador: Prof. Dr. Edson Luiz Ursini - Profa. Dra. Diama Bhadra Vale

Este trabalho corresponde à versão final da Tese defendida por Juan Fernando Galindo Jaramillo e orientada pelo Prof. Dr. Paulo Sérgio Martins Pedro.

Limeira
2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Felipe de Souza Bueno - CRB 8/8577

Galindo Jaramillo, Juan Fernando, 1982-
G133a Agent-based modeling and simulation of social determinants of health / Juan Fernando Galindo Jaramillo. – Limeira, SP : [s.n.], 2023.

Orientador: Paulo Sérgio Martins Pedro.
Coorientadores: Edson Luiz Ursini e Diama Bhadra Andrade Peixoto do Vale.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Simulação (Computadores). 2. Modelagem matemática. 3. Acesso aos serviços de saúde. 4. Neoplasias do colo do útero. I. Martins Pedro, Paulo Sérgio, 1967-2024. II. Ursini, Edson Luiz, 1951-. III. Vale, Diama Bhadra Andrade Peixoto do,1978-. IV. Universidade Estadual de Campinas. Faculdade de Tecnologia. V. Título.

Informações Complementares

Título em outro idioma: Modelagem e simulação basaeada em agentes de determinantes sociais de saúde
Palavras-chave em inglês:
Computer simulation
Mathematical modeling
Health services accessibility
Uterine cervical neoplasms
Área de concentração: Sistemas de Informação e Comunicação
Titulação: Doutor em Tecnologia
Banca examinadora:
Edson Luiz Ursini [Coorientador]
Kaue Tartarotti Nepomuceno Duarte
Joana Fróes Bragança Bastos
Flavio Rubens Massaro Junior
Rosa Cristina Cecche Lintz
Data de defesa: 18-12-2023
Programa de Pós-Graduação: Tecnologia

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-2406-5100
- Currículo Lattes do autor: https://lattes.cnpq.br/3826625800528328

FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da sessão pública de defesa de dissertação para o Título de Doutor em Tecnologia na área de concentração Sistemas de Informação e Comunicação, a que se submeteu o aluno Juan Fernando Galindo Jaramillo, em 18 de dezembro de 2023 na Faculdade de Tecnologia – FT/UNICAMP, em Limeira/SP.

Prof. Dr. Edson Luiz Ursini
Presidente da Comissão Julgadora

Dr. Kaue Tartarotti Nepomuceno Duarte
University of Calgary

Profa. Dra. Joana Fróes Bragança Bastos
FCM/ UNICAMP

Prof. Dr. Flavio Rubens Massaro Junior
Fundação Hermínio Ometto

Profa. Dra. Rosa Cristina Cecche Lintz
FT/UNICAMP

Ata da defesa, assinada pelos membros da Comissão Examinadora, encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-graduação da Faculdade de Tecnologia.

Agradecimentos

Aunque esta tesis lleve solamente mi nombre, hay muchísimas personas cuya contribución es invaluable y que hicieron posible que fuera realizada. En primer lugar, quiero agradecer a todos los profesores que me orientaron durante este proyecto. Paulo Martins me invitó a investigar en Modelos y Simulación Basados en Agentes. Lastimosamente, al momento de escribir esta tesis, él se encuentra en un estado delicado de salud. Espero que pronto se recupere.

A Edson Ursini, quien decidió apoyarme para continuar con este proyecto cuando Paulo tuvo que retirarse para tratar su enfermedad, solo puedo agradecerle. Gracias a su experiencia he podido encontrar el camino para completar esta tesis. Igualmente, a Diama Vale, quien siempre fue una fuente de conocimiento e inspiración, quien creyó en la posibilidad de hacer trabajo interdisciplinar, con las dificultades que este conlleva, y quien siempre me ayudó a direccionar la investigación y a interpretar los resultados desde el punto de vista de la Medicina y la Salud Pública, quiero expresarle mi total gratitud.

Algunas personas estuvieron presentes en la elaboración y revisión de este trabajo y me ayudaron bastante a realizar esta tesis. En ese sentido, a Giovana Formigari, alumna de Medicina de la FCM-UNICAMP, le quiero agradecer toda su disposición para el levantamiento de los datos y su interpretación desde su campo de acción. Estoy seguro de que llegará muy lejos en su carrera.

También quiero agradecer a los miembros de las bancas de calificación y defensa, pues me ayudaron a definir mejor el proyecto de investigación y a mejorarlo en forma y fondo. También quiero agradecer a todos los profesores y personal administrativo de la universidad, quienes de alguna forma contribuyeron a la realización de este proyecto. También quiero agradecer a mis colegas de posgrado, con quienes siempre tuve discusiones que me ayudaron a mejorar mi tesis.

Finalmente, quiero agradecer a mi familia por su paciencia y apoyo. Fueron muchas noches, muchos fines de semana y muchas vacaciones en los que me dediqué a este proyecto. Gracias por todo su apoyo durante este proyecto. A Pamela, Matías y Melo, quiero agradecerles por tantas alegrías que me ayudaron a llevar momentos difíciles durante mi doctorado. A mi mamá, María Isabel y Antonella, gracias por acompañarme en la distancia.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

A modelagem de sistemas complexos é útil para a predição de cenários em que há presença de fenômenos emergentes. O uso de Modelagem e Simulação Baseada em Agentes (ABMS) permite a previsão dos fenômenos emergentes de um sistema, a partir das interações dos agentes do sistema entre si e com o ambiente em que estão inseridos. A análise de políticas públicas requer a identificação, o estudo e a caracterização de fenômenos complexos relacionados com sistemas complexos. No caso dos sistemas com Determinantes Sociais de Saúde (SDoH) é necessário considerar o ambiente social no qual indivíduos e instituições públicas e privadas interagem, e como essas interações impactam a saúde da população. Portanto, o uso de ABMS na modelagem de cenários em saúde pública que consideram fenômenos sociais pode resultar útil na identificação de como diferentes variáveis sociais influenciam a saúde pública. Assim, o uso de ABMS resulta conveniente na tomada de decisão em políticas públicas de saúde. Neste projeto, apresenta-se um método de modelagem de cenários de saúde pública considerando SDoH. Inicialmente, foi criado um modelo de regressão logística com variáveis clínicas e demográficas, a partir de dados da Fundação Oncocentro e do SEADE. Essa regressão permitiu estabelecer uma relação significativa de algumas variáveis com a detecção do câncer de colo de útero em uma etapa inicial. Dentre essas variáveis, destaca-se o Índice Paulista de Responsabilidade Social, um indicador do Estado de São Paulo que toma em consideração escolaridade, riqueza e longevidade na sua construção. O modelo de regressão logística foi utilizado para a calibração do ABMS. Esses modelos foram criados replicando os resultados gerais a partir de testar como a partir da capacidade de atendimento ou na disponibilidade de horários no final de semana, indicando que poderiam ter algum impacto na detecção precoce do câncer de colo de útero. Dessa maneira, a construção de ABMS para a análise do acesso ao rastreamento do câncer permite a tomada de decisão em relação a políticas de prevenção e detecção precoce. Todavia, a avaliação e definição dos cenários testados depende quase inteiramente da intuição do modelador. O uso de técnicas de aprendizagem por reforço pode ajudar a fazer a avaliação de cenários mais sistemática. O método proposto poderá ser utilizado para a identificação de boas estratégias de melhoria das condições sociais que afetam a saúde na esfera pública.

Abstract

Complex system modeling is a valuable tool for predicting scenarios involving emergent phenomena. Agent-based modeling and simulation (ABMS) can be used to forecast these phenomena by simulating the interactions of individual agents within a system and their environment. Public policy analysis requires identifying, studying, and characterizing complex phenomena related to complex systems. In systems with Social Determinants of Health (SDoH), it is crucial to consider the social environment where individuals and public and private institutions interact and how these interactions impact population health. Therefore, ABMS can be a valuable tool for modeling public health scenarios that consider social phenomena to identify how different social variables influence health outcomes. This makes ABMS a valuable tool for public health policy decision-making. This project presents a method for modeling public health scenarios considering SDoH. A logistic regression model was initially developed using clinical and demographic data from the Fundação Oncocentro and SEADE. This regression established a significant relationship between some variables and early detection of cervical cancer. Among these variables, the Índice Paulista de Responsabilidade Social (IPRS) stands out. This indicator, developed by the State of São Paulo, considers education, wealth, and longevity in its construction. The logistic regression model was used to calibrate the ABMS. These models were created by replicating general results and testing how different scenarios, such as changes in service capacity or weekend appointment availability, could impact early detection of cervical cancer. The ABMS constructed for the analysis of access to cancer screening allows decision-making regarding prevention and early detection policies. However, the evaluation and definition of the tested scenarios depend almost entirely on the modeler's intuition. Reinforcement learning techniques can help make scenario evaluation more systematic. The proposed method can be used to identify effective strategies for improving social conditions that impact public health.

List of Figures

List of Tables

List of Acronyms

ABMS	Agent-Based Modeling and Simulation
EBMS	Equation-Based Modeling and Simulation
HDI	Human Development Index
HPV	Human Papillomavirus
ISR	Index of Social Responsibility of the State of São Paulo
SDoH	Social Determinants of Health

Contents

Chapter 1

Introduction

In this chapter, the challenges present in this research are described and an overview of the methodology used is presented.

1.1 Overview

Cervical cancer is the fourth most frequent cancer type among women, responsible for more than 340,000 deaths in 2020, 90% in developing countries (WORLD HEALTH ORGANIZATION, 2022). It is a treatable disease when detected at an early stage, yet most cases in developing countries are detected at an advanced stage (FERLAY et al., 2013). For instance, around 80% of cases in Brazil are detected at an advanced stage (VALE, D. B.; SAUVAGET, et al., 2019). Policymakers must establish strategies to increase the proportion of detection at an early stage to reduce mortality numbers.

Medical screening, that is, testing individuals or populations to detect unknown health conditions, is a widely used practice for cervical cancer detection(WORLD HEALTH ORGANIZATION, 2022). For cervical cancer detection, regular screening is recommended to detect precursor lesions or early cancer and avoid further complications. It is recommended to test women between the of ages 25 and 65 every three years. If there is a positive test, the woman should be referred to a diagnostic test and managed properly (INCA, 2016).

The percentage of coverage of screening in the population is related to the reduction in mortality. It is expected less than three deaths per 100 000 people for coverage above or equal to 50% and less than two deaths per 100 000 people for coverage above or equal to 70% (INCA,

2016). Though, socio-economical conditions are related to access to healthcare in general and cervical cancer detection in particular (TEMKIN et al., 2018).

There is evidence that indicates the relationship between socio-economic status variables, such as Human Development Index (HDI), gross domestic product (GDP) per capita, illiteracy rate and fertility rate, and cervical cancer incidence and mortality (DENNY et al., 2016; VALE, D. et al., 2019). It is also noticeable that socio-economic conditions affect access to healthcare services (TEMKIN et al., 2018). This situation shows the need for considering these indicators in a study to maximize the early detection of cervical cancer.

The Social Determinants of Health (SDoH) are social and economic elements that impact how people live (BUSS; PELLEGRINI FILHO, 2007). As SDoH are helpful for the identification of the relationships between social conditions and diseases, they might be helpful as well for the creation of models to reduce the impact of diseases on a population and, thus, to improve active response in public health policy.

Some studies show SDoH are correlated to cervical cancer incidence and detection (VALE, D. et al., 2019; TEMKIN et al., 2018). Yet, SDoH are usually highly interconnected between them. For example, schooling level may influence income level. Also, policymakers may be interested in knowing different scenarios and the impact of different strategies in these scenarios, creating the need to consider complexity when modeling cervical cancer detection within a population. Those challenges involve anticipation and harnessing of relationships between diverse and changing entities with individual decisions (PAGE, 2010). This shows SDoH may create a complex scenario.

Complex phenomena, studied in different areas, such as physics, biology, and social sciences, are hard to define and quantify (LINEWEAVER; DAVIES; RUSE, 2013). To deal with the lack of consensus on a formal definition, I used two core principles of complexity established by Page (2010): first, the BOAR (complexity lies Between Order and Randomness) principle relates to the intuition that complex outcomes are neither perfectly ordered nor totally random; second, the DEEP (complexity cannot be easily Described, Evolved, Engineered, and Predicted) principle in turn, establishes the difficulties when working with complexity. Considering these principles, it is possible to consider complexity as an outcome from a system hard to predict, yet prone to describe.

Complex systems (CS) are collections of diverse, interconnected, and independent actors that behave according to rules that might adapt according to the environment in which those

actors interact (PAGE, 2010). CS behavior is established by the interactions among its actors. Those interactions allow the system to learn and evolve, making possible the presence of emergent phenomena (FURTADO; SAKOWSKI; TÓVOLLI, 2015).

Thus, in order to deal with complexity, complex systems modeling must identify the emergent phenomena from interactions among individuals present in the modeled system and the effects of the environment in which the system takes place (WILENSKY; RAND, 2015).

One approach for emergent phenomena identification in CS is based on modeling the behavior of the constituent parts of the system and their interactions to determine the overall system response. Different names are found in literature for this approach. It can be called Agent-Based Modeling and Simulation (KLÜGL; BAZZAN, 2012; CONTE; PAOLUCCI, 2014), Agent-Based Modeling (RAILSBACK; GRIMM, 2019), Agent-Based Simulation (KIESLING et al., 2012), Agent-Based Models (GIABBANELLI; CRUTZEN, 2017), Agent-Based Computer Modeling (EPSTEIN; AXTELL, 1996), or Multi-Agent-Based Simulation (SICHMAN, 2015). I will use the name Agent-Based Modeling and Simulation (ABMS) as it is the most broad-ranging.

This approach is based on modeling the interactions between agents - autonomous individuals that have properties and actions - and the environment (WILENSKY; RAND, 2015). ABMS is derived from Thomas Schelling's proposal to obtain emergent patterns from simple spatially distributed models of individual behavior (SCHELLING, 1969). The basic idea of ABMS is shown in Figure 1.1.

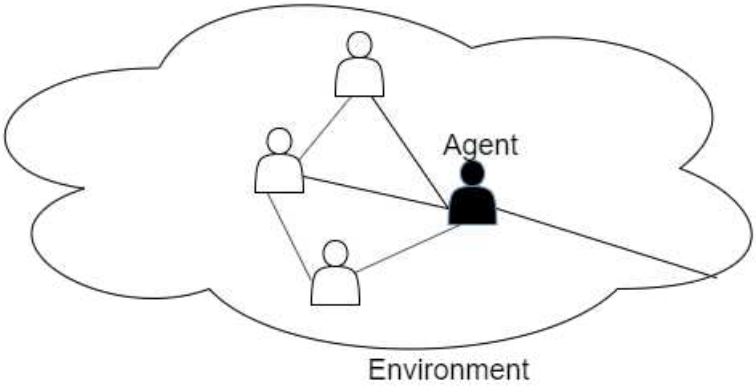


Figure 1.1: Basic ABMS

ABMS has gained popularity during the last decade, being used to model complex systems in areas such as social sciences, combat scenarios, urban planning, food behaviors,

evacuation scenarios, public policy, biology, and education among others (WEIMER; MILLER; HILL, 2016; GIABBANELLI; CRUTZEN, 2017; KLÜGL; BAZZAN, 2012; FURTADO; SAKOWSKI; TÓVOLLI, 2015). In general, ABMS of social processes can be called artificial societies, as they are intended to create social structures from local interactions between agents (EPSTEIN; AXTELL, 1996).

Public policy analysis compares and evaluates alternatives to identify and implement a desired social change (GENTILE, J.; GLAZNER; KOEHLER, 2015; GIABBANELLI; CRUTZEN, 2017). In the context of public policy, modeling interactions among communities, citizens, and public and private institutions is useful to determine the impact of public policies and make better decisions (FURTADO; SAKOWSKI; TÓVOLLI, 2015).

Public policies usually need to consider citizens and organizations behavior to achieve a desirable outcome. As each agent reacts differently, and interactions between them occur, outcomes of public policy are emergent products of individual decisions and the influence of those decisions with each other and with the policy (RAND, 2015).

Public policy analysis is benefitted from the understanding and prediction of the possible emergent phenomena present in social systems. Those systems are composed by autonomous agents who behave with bounded rationality. That behavior and environmental issues make social systems non-deterministic (GENTILE, J.; GLAZNER; KOEHLER, 2015). Then, it is convenient to model each agent, its behavior, and its interactions with other agents and the environment.

In public health policy, besides biological factors, it is necessary to considerate social variables that may influence the emergence and spread of diseases and that may generate inequity in health conditions among individuals or social groups within a society (WHITEHEAD, 1991). In this context, Social Determinants of Health (SDoH) are the conditions in which individuals within a population live and work (BUSS; PELLEGRINI FILHO, 2007). The use of ABMS for public health models shows promising results in identifying how SDoH impact public health policy (ALVAREZ-GALVEZ; SUAREZ-LLEDO, 2019; OH et al., 2020).

As seen previously, ABMS is useful for modeling social phenomena in general, and public health policy in particular. More importantly, ABMS is the only known modeling approach in which heterogeneous actors behaviors and their interactions are modeled (CONTE; PAOLUCCI, 2014). Yet, ABMS has to face challenges related to agent behavior modeling. One

of the challenges is related to the generation of outcomes interpretable and helpful for decision makers.

Policymakers need tools to help them make decisions based on models organizing empirical, theoretical, and expert knowledge about the policy's goal. These type of models, called justified stories, are intended for decision makers to answer their questions about different scenarios and the choices they can make (BADHAM et al., 2021). The value of these models is the possibility of accessing the knowledge available to analyze different strategies in different scenarios. Then, ABMS must include variables, parameters, and outcomes reflecting these scenarios.

Traditional ABMS needs that agent behavior rules are well established before running a simulation (WEIMER; MILLER; HILL, 2016). Those rules are determined in the agent architecture, that is, the software architecture that determines agents' learning and actions (CHIN et al., 2014). Yet, a rule-oriented instead of an agent-oriented approach limits models' scope. This is because in a changing environment with bottom-up and top-down interactions, agents behavior change. A generative approach based on a cognitive agent theory is necessary to make a model in which learning and adaptation are incorporated (CONTE; PAOLUCCI, 2014). That approach must be considered from the agent architecture.

In this thesis, the creation of a method for ABMS modeling is presented. In this method, a logistic regression model is used for validation and calibration. This method should optimize the model outcomes, according to the rewards given to the agents.

1.2 Research Questions

Q: How to systematically test different scenarios for cervical cancer detection at an early stage, considering SDoH and behavior issues?

1.3 Research Hypotheses

H: By using ABMS with SDoH, it is possible to identify the influence of different strategies in early detection of cervical cancer.

1.4 Research Objective

The general objective of this research is to develop a method to systematically model and evaluate public health scenarios related to cervical cancer detection at an early stage, considering SDoH and using ABMS.

According to the general objective, the following specific objectives were defined:

- To create a dataset with SDoH data related to cervical cancer.
- To develop a statistical model for calibration of cervical cancer scenarios.
- To model and simulate in ABMS for cervical cancer detection at an early stage, in which biological and social variables are considered.

1.5 Methodology

In order to test the research hypothesis, the following steps were performed.

The first step is the creation of a dataset by joining cervical data records using data from the São Paulo Oncological Foundation (FOSP) with demographic data published by the State of São Paulo Statistics Portal (SEADE). All variables were reviewed and some pre-processing techniques, such as data normalization, imputation, and selection, were applied (GALINDO; FORMIGARI; VALE, D. B., et al., 2021).

The second step is the creation of a logistic regression (logit) model to select the variables used in the SDoH. The logit model is used for calibration and validation of the ABMS model.

Finally, in the third step is the creation of a baseline ABMS model to check the proportion of cervical cancer detection from different strategies.

The rest of this work is shown as follows: in Chapter 2 the theoretical framework supporting this thesis is presented. In Chapter 3, a literary review about ABMS with SDoH is shown. Chapter 4 shows the Methodology followed to answer the research question. The data preprocessing and the logit regression are presented in Chapter 5, and the base ABMS results are shown in Chapter 6. Finally,conclusions and propose future works are shown in Chapter 7.

Chapter 2

Basic Concepts

In order to clarify some important concepts for the research taking place and determine its boundaries, some basic concepts about SDoH, complexity, modeling, and simulation on public policy are explored in this chapter. Concepts related to ABMS are emphasized, as it is the focus of this research project.

2.1 Cervical Cancer and Social Determinants of Health

Even though cervical cancer causes, prevention, and treatment are well established, it is still the fourth most frequent cancer among women in the world (WORLD HEALTH ORGANIZATION, 2022) and most cases still develop into advanced stages in which treatment is more complicated, painful, and costly (VALE, D. B.; SAUVAGET, et al., 2019). Moreover, there are persistent inequalities in early detection of cervical cancer related to the level of development of each country.

These inequalities are also present in Brazil, as the proportion of cervical cancer stages detected at an early stage is higher in the Southeast Region and lower in the North and Northeast regions (VALE, D. B.; SAUVAGET, et al., 2019). These disparities may be related to regional variables, such as fertility rate and Human Development Index, as well as disparities in access to healthcare services (VALE, D. B.; TEIXEIRA, et al., 2021). Considering this, it is necessary to understand, not only the causes of cervical cancer, but also the social and environmental issues related to its incidence. Also, the study of the incidence of Social Determinants of Health (SDoH) in the proportion of cervical cancer cases at an advanced

stage may suggest this proportion as an index to understand the general health conditions within a population.

Cervical cancer is caused in virtually all cases by the Human Papillomavirus (HPV) (SCHIFFMAN; CASTLE; JERONIMO, et al., 2007a). As it may take decades from HPV infection to the presence of symptomatic cancer, a cervical cancer screening policy is necessary for early detection of cases in apparently healthy women.

The International Federation of Gynecology and Obstetrics (FIGO in French) determined four stages for the extent of cancer (SALIB et al., 2020). They go from I to IV, as follows:

- Stage I: Local tumor
- Stage II: Tumor invades surrounding organs or tissues
- Stage III: Tumor invades distant tissues within the pelvis
- Stage IV: Existence of distant metastasis

Precursor lesions to cancer can be considered stage 0, though they are not considered in the FIGO classification. Stage I corresponds to the early stage, in which the 5-year survival rate goes between 79% and 96%, and treatment is easy and less costly. Stages II or higher correspond to an advanced stage, in which survival rate goes from 70% to 14% and treatment is expensive and painful (SALIB et al., 2020). At an early stage, cervical cancer is asymptomatic and can only be detected though screening.

Effective cervical cancer screening and treatment policies have reduced mortality fivefold in the countries in which it is implemented (WORLD HEALTH ORGANIZATION, 2022). To define screening policies, it is necessary to define the target population, the screening method(s) and the screening periodicity. These considerations are interrelated, as depending on the effectiveness of the screening method in the target population, different screening periods can be applied.

There are different cervical cancer screening methods. Traditional cytological testing (Papanicolau smear) has shown its effectiveness in identifying pre-cancer lesions or cancer at the first stage. HPV testing has shown better results in screening than cytological testing and is the preferred testing method according to the World Health Organization (2022), screening policy in Brazil is still based on cytological testing (INCA, 2016).

Considering this, the Brazilian Ministry of Health recommends testing healthy women in ages between 25 and 64 using the traditional cytological procedure (INCA, 2016). There are different considerations regarding the results. Yet, they will be not considered in this work, as it is focused on public policy as a whole instead of individual outcomes.

A successful screening policy depends on the screening rate. This rate is calculated as the percentage of women in ages between 25 and 64 who live in a given area being tested within the last three years (INCA, 2014). An 80% coverage rate was defined as a target in Brazil (INCA, 2022). Yet, screening numbers have decreased in the last years and differences in access between Brazilian regions are noticeable (SILVA, G. A. et al., 2022). This shows the need to consider equity in cervical cancer screening policies.

In the context of health, equity must be understood as giving to all people the same opportunities to achieve their full health potential (WHITEHEAD, 1991). That definition of equity takes into consideration the existence of biological differences that make some people more prone to develop a particular illness than others. Yet, some differences are related to the conditions in which people live and work, health attention costs, and limitations to healthy habits related to economical or social issues. Thus, in order to design, implement, and evaluate health policy, it is necessary to consider both biological and social variables that may interfere with health for different groups and different geographical locations.

Implementation of health policy requires, then, the inclusion of social variables that affect health within communities. The concept of SDoH represent the efforts in identifying, understanding, and controlling those social variables. Even though there are several definitions on SDoH, all of them relate to living and working conditions of individuals and human groups and the incidence of those conditions on health (BUSS; PELLEGRINI FILHO, 2007). Variables as diverse as access to potable water, distance to the nearest health center or language barriers for immigrants are examples of social conditions that affect health within a population.

The understanding of SDoH, therefore, allows the creation of public policy focused on minimizing the differentials between individuals and social groups to reduce health inequities. It is necessary to consider the relationship between SDoH and macroscopic health indicators. The greatest challenge to establishing that relationship consists in establishing a hierarchy of determinants from general factors and the mechanisms by which those factors influence groups and individuals' health (BUSS; PELLEGRINI FILHO, 2007). Yet, those mechanisms are

not linear and have several feedback relationships, being necessary a systemic approach to these relationships (CAREY; CRAMMOND, 2015). Then, to study the incidence of SDoH on health, it is convenient to understand public health as a complex system.

2.2 Complexity

Cells, ecosystems, a galaxy, the internet, the stock market, and politics have in common the presence of complexity. Complex phenomena are present in both hard sciences and humanities. Yet, it is in human sciences in which they create most challenges. Hard sciences are built upon general, fundamental laws. Those laws are tested by experiments that can be replicated and, given the same initial conditions, those fundamental laws are tested and refined.

Social phenomena are different. First, they are discrete by nature, Also, social phenomena do not follow universal patterns and have to deal with uncertainties related to subjectivity (FURTADO; SAKOWSKI; TÓVOLLI, 2015). Some subjectivity issues present in social sciences are related to bounded rationality. Bounded rationality is the fact that human decisions for complex problems is not fully rational, as it escapes human cognition abilities (ARTHUR, 1994). Then, problems in public policy usually deal with complexity, as every individual in a group reacts differently.

Interactions among individuals are also determinant in public policy. As individuals interact and their reactions influence and feedback each other, interactions make the overall result different from the sum of all individual reactions. That makes the outcome of the implementation an emergent product of individual decisions and the interactions of those decisions between them and with the public policy (RAND, 2015). In this context, emergence is understood as ”stable macroscopic patterns arising from the local interaction of agents” (EPSTEIN; AXTELL, 1996, p. 35).

As seen previously, complexity is highly related to actors individual decisions and the effect of those decisions on the system. Then, it is necessary to establish what constitutes a complex system. There are multiple complex systems (CS) definitions, most of them focused on the knowledge area or the particular characteristics of the model. Yet, it is possible to identify some features most definitions have in common (FURTADO; SAKOWSKI; TÓVOLLI, 2015; RAND, 2015; WILENSKY; RAND, 2015; FUENTES, 2015; SICHMAN, 2015; PAGE, 2010):

- Multiple interactions among constituent parts of a system, from and across scales, making it impossible to describe that system considering only the attributes of its parts (RAND, 2015; PAGE, 2010).
- Multiple abstraction levels, from individuals to organizations.
- Self-organization of the system without the need of central control. For example, bird flocks form complex patterns without central coordination (WILENSKY, 1998).
- Presence of feedback. Interactions among parts of the system have effects in time, making the system adaptive and evolutionary.
- Individuals' behavior is determined by rules and may be adaptive (PAGE, 2010).

Rand (2015) identifies some properties that characterize complex systems:

- Emergence: An emergent property is a property of a system that cannot be neither understood nor predicted from the analysis of its individual agents. It results from the interactions of the agents between them and with the environment.
- Leverage points: A place of a CS on which the system can be drastically altered or changed.
- Tipping points: A point where a system suddenly changes states after a small change of one parameter.
- Path dependence: Degree of dependence of the outcome of the system on the initial conditions.
- Nonlinearity: Inputs do not affect outputs linearly.
- Robustness: How much a system maintains its behavior after a perturbation.
- Diversity and heterogeneity: Refer to the uniqueness of individuals. Complexity is also a result of diversity and heterogeneity, as individuals react differently to interactions.
- Interconnectedness and interactions: Connection patterns between individuals can determine the outcome of the system (BARABÁSI; ALBERT, 1999; GAO; BARZEL; BARABÁSI, 2016).

From those properties, it is possible to consider complexity, for the purpose of this work, as an interesting outcome produced from a system composed by the interaction of its parts (PAGE, 2010). Therefore, to deal with complexity, it is necessary to understand the underlying systems that produce it.

2.3 Complex System Methods

CS features and properties make it impossible for a predefined algorithm to predict its outcomes (SICHMAN, 2015). Then, it is necessary to use CS modeling tools being able to explain and identify emergence properties from heterogeneity and diversity (RAND et al., 2003). Diversity also increases robustness and drives innovation and productivity to CS (PAGE, 2010), being desirable, yet hard to deal with.

To deal with CS, some computational and interdisciplinary methods have been proposed. CS methods focus on identifying and predicting adaptive and evolutionary behavior. In these predictions, it is important to understand how small individual decisions, in most cases unaware of the global impact of those decisions, lead to fit and optimal responses for a community (MITCHELL, 2011). This is necessary to understand how individuals adapt to environmental changes, such as the ones made by public policy, and to see how their response affect the overall outcome (RAND, 2015).

There are several methods proposed to model CS. They all consider the characteristics of complexity, emphasizing on different aspects and taking continuous or discrete approaches (FUENTES, 2015). Yet, they differ on their focus and on their nature, making the decision on which method to use depending on the focus of the study of the CS. Table 2.1 shows some of the most used CS methods.

Top-down approaches like macro-simulations based on mathematical models and statistics are appropriate when central control is determinant or in which population individual characteristics are homogeneous, being possible to average population behavior. In opposition, bottom-up approaches, like cellular automata and ABMS have high degree of localization and distribution. They are appropriate for models in which, even existing central control, any component can influence the system.

Table 2.1: Most Used Complex Systems Methods

Method	**Principle**	**Nature**
Nonlinear Science	Nonlinear Equations	Top-down
Bifurcation Theory	Structural Changes in Differential Equations	Top-down
Network Theory	Connections between elements of a system	Bottom-up
Game Theory	Statistics	Top-down
Information Theory	Statistics	Top-down
System Dynamics	Differential equations paired with causal loop diagramming	Top-down
Cellular Automata	Rules mediated	Bottom-up
Agent-Based Modeling and Simulation	Rules mediated	Bottom-up
Data Mining	Rules, Statistics, Differential Geometry	Top-down / Bottom-up

Modeling is broadly used to predict and test different scenarios in public policy (GIABBANELLI; CRUTZEN, 2017). Modeling approaches can be grouped into three categories:

- Qualitative aggregate models
- Quantitative aggregate models (macrosimulation), like differential equations and macroeconomics.
- Quantitative individual models (microsimulation), like network modeling and ABMS.

ABMS is usually defined in opposition to Equation-Based Modeling and Simulation (EBMS) (CONTE; PAOLUCCI, 2014). Because of its nature, ABMS differs from EBMS in that ABMS is able to represent heteronomous populations in discrete space and time and considering individual characteristics (EPSTEIN; AXTELL, 1996). Other approaches divide individuals into groups of homogeneous characteristics (Differential Equations) or use representative agents (macroeconomics). Data mining approaches also consider heteronomy by working with outliers within the systems. Yet, ABMS describes and tests causal relationships from model designing stages, whereas data mining only discovers and tests correlations (GENTILE, J.; GLAZNER; KOEHLER, 2015).

Another difference between ABMS and EBMS is its focus on dynamics. Mathematical social science approaches like macroeconomics tend to focus on equilibrium states. Differential Equations approaches are able to represent dynamics, yet they separate transient and equilibrium states. Finally, ABMS is a top-down approach, differently from most EBMS that take bottom-up approaches (RAND, 2015). That approach also reflects a difference in the understanding of simulations as problem-solving techniques or as descriptions of patterns of behavior (DURÁN, 2018).

2.4 Agent-based Modeling and Simulation

ABMS is a bottom-up approach that models emergent phenomena from individual agent behavior. ABMS consider boundedly rational agents, including outliers and their effects on the system, something most statistical approaches are not able to do (GENTILE, J.; GLAZNER; KOEHLER, 2015). Because of its descriptive capability, ABMS captures individual decision-making processes, as well as the macro effect of interactions between individuals and the environment (GIABBANELLI; CRUTZEN, 2017).

As an interdisciplinary approach in social sciences, ABMS can be described as part of a broader area that studies agents' computational technology to simulate social phenomena, called Agent-Based Social Simulation (ABSS). Sichman (2015) establishes ABSS as the intersection of three disciplines. ABMS is composed by Computer Simulation, and Agent-Based Computing — focused on the study, design, and implementation of artificial agents. The Venn Diagram of those three disciplines is shown in Figure 2.1.

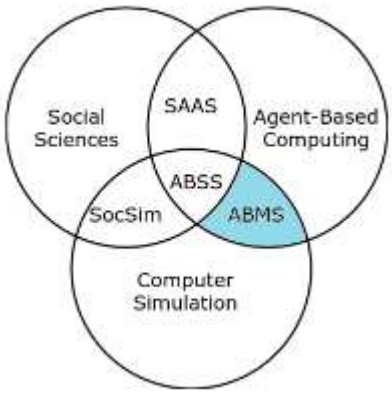


Figure 2.1: ABMS Interdisciplinary Venn Diagram. Source: Adapted from Sichman et al. (2015)

Besides ABMS and ABSS, other elements appear as interdisciplinary intersections in Figure 2.1. Social Aspects of Agent Systems (SAAS) is the study of social constructs present in societies that can be used as an inspiration for computational models used in the analysis and implementation of norms, institutions, and other social relationships. Social Simulation (SocSim), in turn, is the study of computer simulations of social phenomena. Then, ABSS can be understood as the intersection of ABMS with SocSim or SAAS, or as the application of ABMS in Social Sciences. Another view from Conte and Paolucci (2014) puts ABMS as an intersection between Agent Theory, Systems and Architectures and Social Sciences. Both approaches coincide in putting ABMS as an interdisciplinary approach in which computational resources are used to model and simulate social phenomena.

From the interdisciplinary relationships previously presented, it is possible to conclude that ABMS is prone to simulate emergent social phenomena from individual agent behavior descriptive models. ABMS is used mainly for artificial social models to test ideas, and for socio-cognitive and social models to simulate system outcomes from individual cognitive decisions. By simulating those models, it is possible to prototype public policies and analyze different scenarios (DAVID et al., 2004). Then, ABMS is used for comprehension of social complex phenomena, for helping stakeholders to make decisions from predictive simulations, and even for participatory simulations among stakeholders in a common complex environment (SICHMAN, 2015).

As said before, ABMS is focused on modeling agents' individual behavior and their interactions among them and with the environment in order to analyze emergent phenomena (RAILSBACK; GRIMM, 2019). To understand the principles of ABMS, it is necessary to consider some basic definitions.

To begin, it is necessary to establish what constitutes an agent. An agent is an autonomous entity that operates in transitions between states of the world based on mechanisms and representations somehow incorporated into them (CONTE; PAOLUCCI, 2014). The degree of autonomy is important to establish the use of agents. Most models use agents in a "weak" sense, that is, they are not able to sense, manipulate, and reason in their environment, being limited to follow the rules they were programmed to. In opposition, when an agent has sensing, reaction, and learning capabilities, it is an agent in a "strong" sense (CONTE; PAOLUCCI, 2014).

Symbolic representations of agents and environment characteristics are used to allow an agent to actively interact and take decisions. The environment is the medium in which the agents are inserted in (RAILSBACK; GRIMM, 2019). The environment has its own properties. Those properties usually vary in space and time.

Agents have a relationship full of feedback loops with the environment, in which, agents' actions may affect the environment and the environment, in turn, affects agents. In order to understand those relationships, and the emergent phenomena they produce, ABMS considers both individual and systemic behavior, connecting different disciplines in one model (EPSTEIN; AXTELL, 1996), like psychology (individual) and sociology (systemic), or biology (individual) and epidemiology (systemic).

Time in ABMS is divided into steps. In each time step, an iteration occurs. During that iteration, all interactions between agents and with the environment occur. Then, time steps represent the timescale in which interactions happen (seconds, hours, days, years, etc.). Then, the definition of a timescale indicates the speed in which those changes happen (RAILSBACK; GRIMM, 2019).

Interactions between agents and the environment on a large scale generate emergent phenomena. Yet, that phenomena is usually non-deterministic. As described earlier, social systems have stochastic behavior. Open and stochastic systems are not well described by deductive reasoning (GENTILE, J.; GLAZNER; KOEHLER, 2015). ABMS models require abduction logic. It is necessary to execute the model multiple times. Each execution is a deduction of a particular scenario, and its outcomes correspond to that scenario. Small changes in parameter values may lead to totally different outcomes. Therefore, it is necessary to execute a model multiple times considering different parameters to cover multiple scenarios. Then, from those multiple executions it is possible to produce and test multiple hypotheses, and infer the best explanation.

By summarizing the outcomes from ABMS executions, it is possible to analyze the results of public policy. Outcomes from ABMS in public policy can be categorized into three categories when compared to expectations (GENTILE, J.; GLAZNER; KOEHLER, 2015; KOEHLER; BARRY; MEYER, 2006). This categorization is relevant for decision-making in public policy The first category is expected valid outcomes, that is, expected input values create desired outcomes. The second category is expected invalid outcomes, that is, expected input values create undesired outcomes. Finally, the third category is unexpected outcomes,

and is in those outcomes in which insight about a system is created (KOEHLER; BARRY; MEYER, 2006). By using abduction logic, comprehension of the modeled system increases and hypotheses are refined and tested.

From ABMS outcomes, public managers can take decisions that support their intended outcomes. Yet, they need to be conscious of the empirical relevance of the model they used. Model relevance should be defined considering the level of detail and the fidelity to a referent. That referent can be real data or a previously tested model used as reference. Axtell defined four levels of empirical relevance for modeling and simulation (AXTELL, 2005; KOEHLER; BARRY; MEYER, 2006):

- Level 0: Micro-level qualitative correspondence-agents that behave plausibly for a given system.
- Level 1: Macro-level qualitative correspondence to the referent.
- Level 2: Macro-level quantitative correspondence to the referent.
- Level 3: Micro-level quantitative correspondence-agents that behave identically to real world.

To better understand these levels, let's see a flocking bird example present in Figure 2.2. In this example, the modeled phenomenon is related to how, as each bird follows the closest bird to it, the flock creates a determined pattern, being this pattern an emergent phenomenon (STONEDAHL; WILENSKY, 2011). In a model with level 0 relevance, qualitative correspondence in individual movements is ensured, that is, each bird follows the bird closest to it. Going into Level 1 relevance, the birds are creating a pattern that somewhat resembles patterns found in real life. For Level 2 relevance, it is necessary to have quantitative parameters, such as velocity, time, or distance, similar to the referent, for the flock as a whole. Finally, for Level 3 relevance, each bird behavior should result in similar parameters found in real life.

Level 0 models correspond to thought experiments. ABMS should minimally achieve Level 1, and preferably Level 2. Level 3 may be considered ideal, though, as the goal of ABMS and to modeling in general is not to reproduce reality exactly as it is but to represent, describe and predict particular phenomena (RAND, 2015), the cost of achieving Level 3 is seldom worthy (GENTILE, J.; GLAZNER; KOEHLER, 2015).

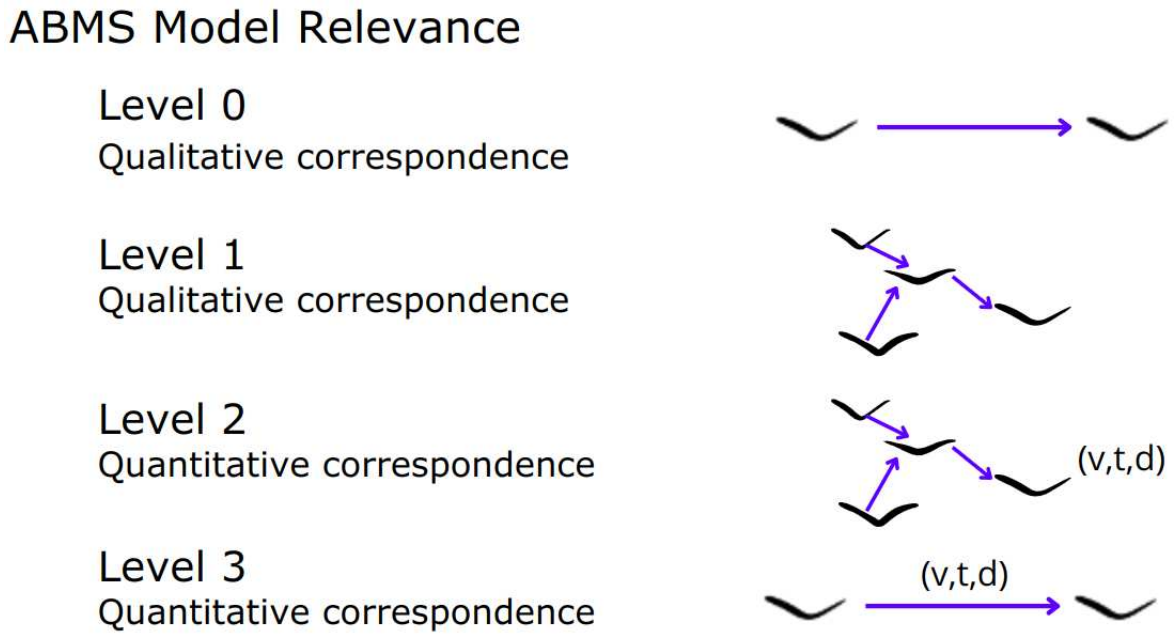


Figure 2.2: ABMS Relevance Model example. Source: Adapted from (AXTELL, 2005)

Axtell et al. (1996) proposed a Docking Framework composed by three levels of correspondence to define the validity of a model, comparing it to a model of reference:

- Identity: identical results to the referent.
- Distributional: Statistically indistinguishable results from the referent.
- Relational: Statistically distinguishable results from the referent, yet qualitatively similar.

For thought experiments, relational level is enough. For decision-making, in turn, it is necessary to achieve at least a distributional level (GENTILE, J.; GLAZNER; KOEHLER, 2015).

Modeling is an iterative process. In each iteration, model relevance and correspondence are improved. Gentile et al. (2015) propose an iterative process, presented in Figure 2.3. That process begins by defining the research question from real world observations. From that research question, a conceptual model is defined. A simulation is built to test that model. To test the simulation and the model, two subprocesses are made. Verification, which compares the simulation with the conceptual model to test its internal validity, and validation, which compares the simulation and external data and checks if the research question is satisfactorily answered, testing the model's external validity.

Model validation consists on testing how good a model represents the system being modeled. In other words, a valid model is a model that correctly satisfies the modeling

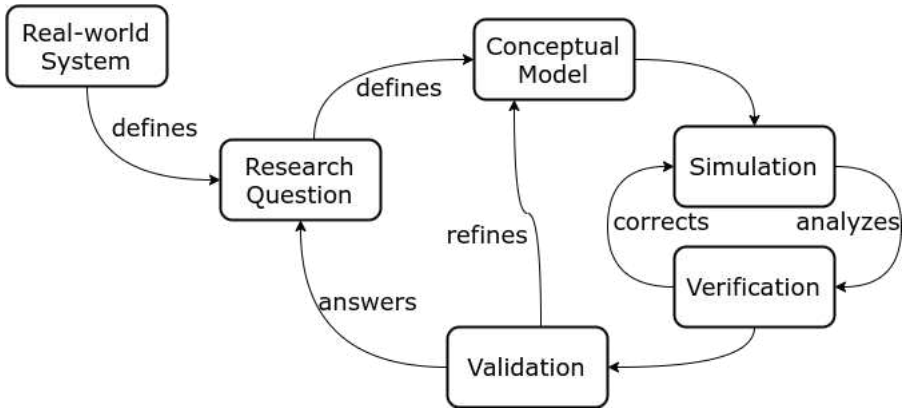


Figure 2.3: Modeling Iteration. Source: Adapted from Gentile et al. (2015)

questions. Because of initial conditions and path dependence, validation is critical in CS. Particularly in ABMS, agents adaptability make model validation harder, as replication of all micro-details present in a model is hardly impossible (GENTILE, J. E.; DAVIS, G. J.; RUND, 2012).

Rand et al. (2003) propose an approach to ABMS validation focused on matching model components and processes to real-world components and processes and matching macro-level patterns, statistics and dynamics found across a variety of cases. Validation can be performed through linear regression using modeling data and comparing it with real-world data. When no real-world data is available, validation can be done by comparing overall behavior of the model with well-known phenomena found in literature.

In most ABMS research, the emergent phenomena are unknown. Also, exploratory research usually lacks of access to real-world data, being impossible to validate the proposed ABMS (GENTILE, J. E.; DAVIS, G. J.; RUND, 2012). A step before model validation is called model verification, in which the simulation is aligned with the conceptual model. Once the conceptual model is considered valid, it is necessary to guarantee that implementation and simulation correspond to that model.

A first step in model verification consists of visually inspecting the inputs and outcomes and compare results with expectation according to the conceptual model. Yet, that type of verification depends entirely on researchers' expertise and ability to detect any deviation. To make verification more systematic, the use of steady-state analysis techniques is useful (GENTILE, J. E.; DAVIS, G. J.; RUND, 2012). Steady-state analysis is made by analyzing the population or distribution of agents in the simulation and parameter calibration. For oscillating systems, Fourier analysis can be done. Yet, this type of verification needs model

behavior to have a steady-state, that is, constant or oscillating behavior. That condition is not always achieved, especially in larger systems. To deal with models for larger systems, Gentile et al. (2015) propose to disassemble the model into components and verify each component. This process is called unit testing. When components are reassembled, it is possible to validate a portion of the system by bringing it to known steady-state phases.

According to Cooley and Solano (2011), there are three model validation stages: model verification (the code does what it is intended to do), model validation (the model reflects the phenomena) and sensitivity analysis (the model is robust enough to accept parameter changes).

Sensitivity analysis is performed by varying initial conditions and parameters and check how these changes affect the outcomes of the model. In most cases, as randomness is added in the construction of the model, sensitivity analysis can be done by running the model several times varying these conditions. Yet, as there are lots of non-linearities, multiple levels of interactions, and emergent properties, and as the mathematical model in most ABMS is unknown, the One-factor-at-a-time (OFAT) method is recommended, as it shows the effect of individual parameter variations in the model and the robustness of the model to these changes, even though, it is not as good as variance sensitivity analysis methods to robustness of patterns (TEN BROEKE; VAN VOORN; LIGTENBERG, 2016).

Cooley and Solano (2011) also make some considerations on the limitations on model validation. The most important is the fact that both the model outputs and real data outputs are stochastic, making exact correspondence virtually unattainable. Another factors, such as path-dependence on ABMS, and the difficulties on the model to reproduce some phenomena, are also highlighted by the authors.

2.5 ABMS Advantages and Limitations

Because of its nature, ABMS have several advantages. First, compared to EBMS and other approaches, ABMS greatest advantage is related to its nature, as it is the only approach in which states of heterogeneous agents, their decisions, interactions, and communications are modeled (CONTE; PAOLUCCI, 2014).

The possibility to represent agents' behavior and connection eases the expression of experimental hypotheses at individual label, making systems easier to model, implement and

visualize. The possibility of considering different types of agents eases separation of processes and entities, as well allows the evaluation and explanation of the effects of individual behavior in simulation final outcomes (RAND et al., 2003).

The establishment of interaction rules is also simple, easing modeling of complex patterns in higher levels. Also, as models themselves are experimental objects, ABMS increases the understanding of those models, making them more prone to inspection and refinement. Those advantages allow the creation of artificial agent societies, in which different hypotheses can be tested, macro-level phenomena can be discovered and characterized, and unforeseen effects can be detected (CONTE; PAOLUCCI, 2014).

The greatest advantage of ABMS is the possibility of modeling both individual behavior and interactions to allow representation, analysis, and prediction of CS for which modeling from equations results virtually impossible (RAILSBACK; GRIMM, 2019). Also, ABMS multi-realizability makes ABMS models generate a higher-level effect from multiple paths (CONTE; PAOLUCCI, 2014). Those advantages create the possibility to create and evaluate more scenarios than EBMS.

All those advantages make ABMS a useful tool for modeling social phenomena, able for models considering SDoH. Yet, there are some limitations in ABMS. The first one is the high computational cost resulting from simulations and data analysis. Designing large-scale ABMS and real-time big data simulations with ABMS require processing of huge amounts of data (CONTE; PAOLUCCI, 2014).

Another limitation, not just for ABMS, but for all CS techniques, is the presence of numerous free parameters. This large number of parameters make model reproduction harder (SICHMAN, 2015). Those parameters are related to path dependence, making is necessary to ensure that model outcomes are robust enough to a wide parameter value distribution (RAND, 2015).

Parameter determination and other modeling tasks need to use knowledge about the topics covered. Though, another limitation is the lack of strong knowledge in the individual level, required to consider heterogeneity and diversity. Sometimes, individual-level data is not necessary, as modeling can be done from theories about individual behavior (RAND, 2015). Another feasible approach can be done by using machine learning, yet it needs a fine-tuned data acquisition strategy.

In order to deal with high computational cost and parameter adjusting, some authors recommend to keep models simple and focus modeling on the relationships of interest for a particular study (WILENSKY; RAND, 2015) Simplify a model surely makes sense, yet it comes at a cost: in order to simplify a model, it is necessary to ignore some elements present in reality, making the model less accurate.

Dealing with simplicity and accuracy is one of the hardest decisions in ABMS. The usual trend is to emphasize in a particular aspect when modeling and to sacrifice those parts that are not considered relevant for the study (WILENSKY; RAND, 2015). Yet, making a model too adjusted for a particular situation may affect the understanding of the CS studied (GIABBANELLI; CRUTZEN, 2017). Also, the higher the interactions in different levels of abstraction, the harder to understand the model (SICHMAN, 2015).

Oversimplifying a model may lead to model inconsistencies as well. Minimality in modeling happens when only the minimum micro-level rules to obtain a known macro-level effect are added to the model. Minimality happens when models are built using backward engineering from known effects, modeling is rule-oriented rather than agent-oriented, model rules are ***ad-hoc***, or when the overall model is inspired by the minimal conditions logic (CONTE; PAOLUCCI, 2014).

Modeling minimality reduces ABMS validity with two possible outcomes. The first one is theory-based ABMS resulting in agent models that distort real agent behavior. The other possible outcome is arbitrary models from ***ad-hoc*** rules, only valid for a tightly limited number of cases (CONTE; PAOLUCCI, 2014).

In order to deal with this validity issue, models based on cognitive theories of agency may be applied to ABMS. Yet, using cognitive agent models such as BDI architectures or models using neural networks increase model complexity, making it harder to guarantee inner validity and calibration.

Even though cognitive ABMS has to deal with increased validity and calibration issues, cognitive modeling is a generative behavioral theory, that is, it describes social phenomena in terms of the external —environmental— and internal —behavioral— mechanisms of the given CS (CONTE; PAOLUCCI, 2014). Generative models allow modeling adaptation to change in environmental conditions or in individual preferences, as the general mechanisms are considered from the beginning. A generative explanation is necessary for modeling

complex social dynamics, in which emergent phenomena retro-act and agents react to environmental changes (CONTE; PAOLUCCI, 2014).

Parameter determination and validation difficulties may affect model readability and repeatability. The ODD (overview, design concepts, and details) protocol was created to improve ABMS documentation readability and repeatability. (GRIMM; BERGER; BASTIANSEN, et al., 2006).

The ODD protocol consists of three blocks, each block composed of different elements, as presented in Figure 2.4. The first block, Overview, gives information about the purpose, the state variables and scales, and process overview and scheduling. The second block, Design Concepts, provides information about the basic principles, emergence, adaptation, objectives, learning, prediction, sensing, interaction, stochasticity, collectives, and observation of the model. Finally, the third block, Details, gives information about initialization, input data, and existing submodels.

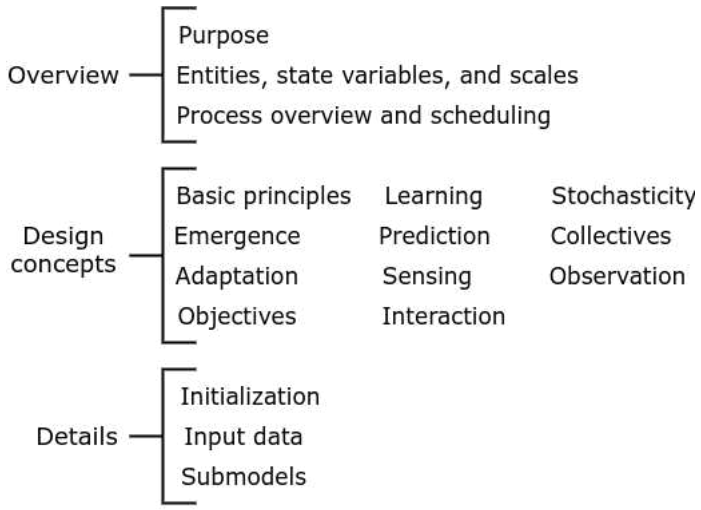


Figure 2.4: ODD Protocol. Source: Adapted from Grimm et al. (2020)

The original protocol was first updated in 2010 to overcome some confusion related to terminology (GRIMM; BERGER; DEANGELIS, et al., 2010; GRIMM; POLHILL; TOUZA, 2017). Yet, the 2010 version of the ODD protocol still had some issues (GRIMM; RAILSBACK, et al., 2020). It was not simple to use, as modelers were not required to describe the structure and processes of the model implementation. Model rationality also was not explicitly required, reducing model credibility. Another issues were the difficulty to reuse an existing model and the tendency to create long documents when using the protocol.

To solve those problems by improving clarity, model repeatability, and structural realism, a new version of the protocol was created (RAILSBACK; GRIMM, 2019; GRIMM; RAILSBACK, et al., 2020). The solution to those issues comes in the form of supplements to ODD. Those supplements, ease ABMS documentation in order to increase model readability and repeatability.

As presented before, two of the greatest advantages of ABMS are the possibility of considering individual behaviors and interactions between individuals and organizations, and ABMS multi-realizability, allowing the analysis of multiple scenarios. Those advantages allow testing several strategies and comparing the results of these strategies, as a support tool for decision-making. However, the process of defining and testing different strategies done manually is limited (CONTE; PAOLUCCI, 2014).

Chapter 3

Bibliographic Reviews

For the purpose of this research, a bibliographic review focused on the use of ABMS including SDoH for public health policy was performed.

3.1 Use of ABMS including SDoH for public health policy

A narrative review of ABMS works including SDoH in public health from 2017 to 2022 using PubMed, Scopus, and Web of Science databases was performed. The five-year interval was defined as a filter to get recent works only. We found 272 results containing agent-based modeling and social determinants of health, considering both journal and conference papers in some part of the text. Yet, when reducing to title and abstract, the number of papers went to 16, focusing on AIDS/HIV, COVID-19, food behavior, and access to healthcare, as shown on Table 3.1. The process can be summarized as follows:

- Search for: (Agent-Based Models) OR (Agent Based Models) OR (Agent-Based Modeling OR Agent Based Modeling) OR (Agent-Based Simulation) OR (Agent Based Simulation) AND (Social Determinants of Health) OR (SDoH).
- Limit the search for years between 2017 and 2022.
- Limit the search for terms appearing in the Abstract.
- Eliminate works that correspond to a previous stage of the same research.

We found that interest in using ABMS to consider SDoH for analyzing different scenarios is quite big, even though, there are few examples in the area. Moreover, there were no results

Table 3.1: Main topic of papers regarding ABMS and SDoH

Main Topic	Conference	Journal	Total
General	2	2	4
COVID-19	2	1	3
HIV/AIDS	1	2	3
Food behavior	0	3	3
Access to healthcare	2	1	3
Total	7	9	16

related to cervical cancer. To explain this lack of novel results publishing, we detected some concerns in model documentation, validation, and reproducibility.

From these 16 works, eleven were chosen by choosing the most recent work from the ones detected.

Tracy, Cerdá, and Keyes (2018) made a review paper to identify the level of adoption of ABMS in public health. Most ABMS are focused on infectious disease epidemiology, as this area is one of the first in which computational models were applied. ABMS has also been applied in social epidemiology to understand phenomena like urban violence or social segregation. ABMS use is also growing in noncommunicable (noninfectious) disease control, for medical conditions such as diabetes or obesity in children and minorities. Another area in which ABMS has also been used is the analysis of health behaviors that increase disease risk, like smoking, sedentarism, alcohol consumption, and unhealthy eating habits.

That review showed ABMS growth in popularity is related to its capacity to give insight for public health policies. Yet, the use of ABMS in public health needs to deal with reproducibility and validation. The authors pointed out the use of the ODD protocol in ABMS for public health to increase reproducibility and use of empirical data for model validation and parametrization.

Another review explored two limitations in ABMS of food behaviors (GIABBANELLI; CRUTZEN, 2017). The first one is the lack of details about the roles played by peers and environment in the models. The second one is the lack of expertise using large amounts of data. To overcome both limitations, the authors of that review recommend an interdisciplinary approach and the use of machine learning.

An example of ABMS of food behaviors is the work of Langellier, Lê-Scherban, and Purtle (2017). The authors built an ABMS to study the effects of the creation of a sugary drink tax on pre-kindergarten attendance, educational achievement and sugar-sweetened drinks consumption in children. Their model, built upon a geographic information system (GIS) of

Philadelphia, allowed the researchers to measure the impact of tax in different populations of children, according to characteristics such as age, gender, race, or the income of their families.

Another example of ABMS of food behaviors is a study to identify the spread of healthy eating habits and its implications on hypertension (KHADEMI et al., 2018). The authors of that study focused on agents' preferences for fruits and vegetables based on their initial preferences, social pressure, and individual characteristics, such as age or gender. The authors used probit regression to estimate the parameters for taste and healthy food preferences.

The work done by Hogan, Galai, and Davis (2021) analyzes the impact of different modeling approaches for HIV incidence and prevalence using SDoH. This paper indicates the importance of ABMS for detecting global trends from individual variables. Yet, it indicates the limitations of this approach depending on the availability of data.

A remarkable work related to HIV/AIDS is the one Rasella and others are currently performing (RASELLA et al., 2022). These authors are analyzing the impact of the Bolsa Família Program and the Family Health Strategy on HIV incidence and treatment in Brazil. They are using data from 2000 to 2018 on HIV/AIDS incidence to perform Regression Discontinuity Design (RDD), Random Administrative Delays (RAD) and Propensity Score Matching (PSM), combined with multivariable Poisson regressions for cohort analyses. With these results, they are planning to perform ABMS to evaluate policies to prevent and treat HIV/AIDS.

A review made by Morshed et al. (2019) found 38 papers on ABMS and System Dynamics. The authors found that most works focused on social network-based influences on obesity, physiology and disease state mechanics, and how food and physical activity environments influence obesity. The authors identified limitations in synthesizing scientific knowledge using the current models due to differences in both temporal and geographical scale, as well as variability both in calibration and validation of models. Those limitations reduce models' realism and validity. Also, lack of documentation affected repeatability of most models. In the same line, Vermeer and others (2022), recommend the use of good practices to encourage modeling validation with local data when possible, as well as making model documentation available to support recommendation and increase adoption of model-based decision-making.

Related to access to health services, two works stand out. In the first one, an ABMS was created to study the mechanisms of reproduction of health inequalities (ALVAREZ-GALVEZ; SUAREZ-LLEDO, 2019). The authors used neural networks to calibrate the model In the second one, the authors studied access to healthcare services in the Latino community of the United States (OH et al., 2020). They made an ABMS in which a percentage of Latinos from a population of 10.000 got sick and needed medical attention. In their model, the authors studied the effects of having Spanish-speaking health professionals and English proficiency in patients in improving medical attention. Regarding the spread of diseases, a work done by Starr and Kain (2022) tested the effectiveness of three policies, mask mandate, testing and isolating, and lockdown, to control the spread of COVID-19 at a municipal level in the United States. The SDoH considered for this work were access to vaccination, population density and size, racial profile, and age demographics. Starr and Kain found that there were different effectiveness depending on age demographics, vaccination rates, and the proportion of Black and Latino populations. This work showed how municipal SDoH need to be considered for the definition of health policies.

From the literature, it is possible to see that ABMS is gaining popularity in different fields of public health policy studies. The inclusion of SDoH in those ABMS is present in several studies. Yet, there are some concerns in validation, reproducibility, and generalization of those studies.

Chapter 4

Proposed Approach

In this chapter, the approach performed in this thesis is described, including the dataset search, pré-processing, imputation, and selection, the creation of a logistic regression model for validation, and the creation, validation, and refinement of the ABMS>

4.1 Proposed Method

In order to achieve the objectives of this research, the following methodology was performed:

1. Searching for datasets with cervical cancer and demographic data.
2. Data pre-processing.
3. Building of a logistic regression as a basis for ABMS validation.
4. Construction of a ABMS according to literature and secondary or primary data (if necessary). Simulation of the model with agent ruled behavior.
5. Verification and validation of the ABMS model with the logit model.

For the first step, a dataset was created by joining cervical cancer incidence data records from Oncocentro Foundation (FOSP), available at http://www.fosp.saude.sp.gov.br/fosp/diretoria-adjunta-de-informacao-e-epidemiologia/rhc-registro-hospitalar-de-cancer/banco-de-dados-do-rhc/ (accessed on March 13, 2022), and cities' demographic records from State of São Paulo Statistics Portal (SEADE), available at

http://catalogo.governoaberto.sp.gov.br/dataset/20-indice-paulista-de-responsabilidade-social-iprs, (accessed on March 13, 2022).

We chose the FOSP dataset, as it is a reliable dataset for cancer in the State of São Paulo. The data from SEADE also includes the Index of Social Responsibility of the State of São Paulo (ISR), a social index whose creation is based on the HDI, but considering local realities (SEADE, 2016).

ISR is an ordinal index, going from level 1 to level 5. Figure 4.1 shows how the ISR is defined, based on city's wellness, longevity and schooling. For level 1 (vulnerable), all wellness, longevity, and schooling are on a low level. A Low level in wellness and a low level in longevity and schooling results in level 2 (in transition). Low wellness and medium longevity and schooling results in level 3 (equitable). High wellness and low longevity or low schooling results in level 4 (unequal). Finally, high wellness, longevity and schooling results in level 5 (dynamic).

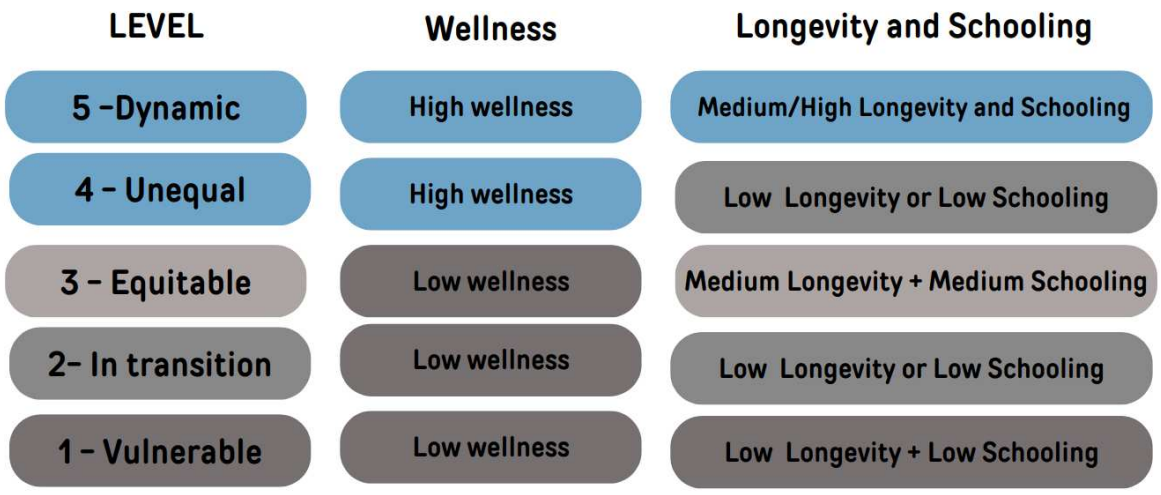


Figure 4.1: Index of Social Responsibility of São Paulo. Source: Adapted from (SEADE, 2016)

After the creation and descriptive analysis of the data, data pre-processing techniques were required. It was necessary to perform data normalization, data imputation, and data selection (GALINDO; FORMIGARI; VALE, D. B., et al., 2021).

With the refined data, we performed a logistic regression using R v4.2.2. For this regression, the dependent variable was the patient's cervical cancer stage, and the independent variables, the other pre-processed and selected data corresponding. More details about the first three items are described in Chapter 5.

For steps 4 and 5, we considered the fact that different types of agents usually have different goals and fitness functions, and also agents goals not necessarily correspond to the desired global outcomes. For that reason, in this work, we first built a ABMS. Then, we validated the model and worked on defining how to expand it. In future models, it may be possible to incorporate machine learning algorithms.

This project was approved by the 'Ethics and Research Committee of UNICAMP, under the number CAAE: 42657020.1.0000.5404. According to the Committee, there was no need for informed consent, as the data used is secondary and patients' identification could not be accessed.

4.2 ABMS Environment

For ABMS coding, implementation, and testing, we used Netlogo v6.2.2. Netlogo is both an open source programming language and a multi-agent programmable environment to ease ABMS design, programming, testing, and running (TISUE; WILENSKY, 2004). Netlogo environment, shown in Figure 4.2 is composed by three tabs: the Interface tab, selected in Figure 4.2, in which the user interacts with the model; the Info tab, containing the model's documentation; and the Code tab, which contains the underlying code used in the model.

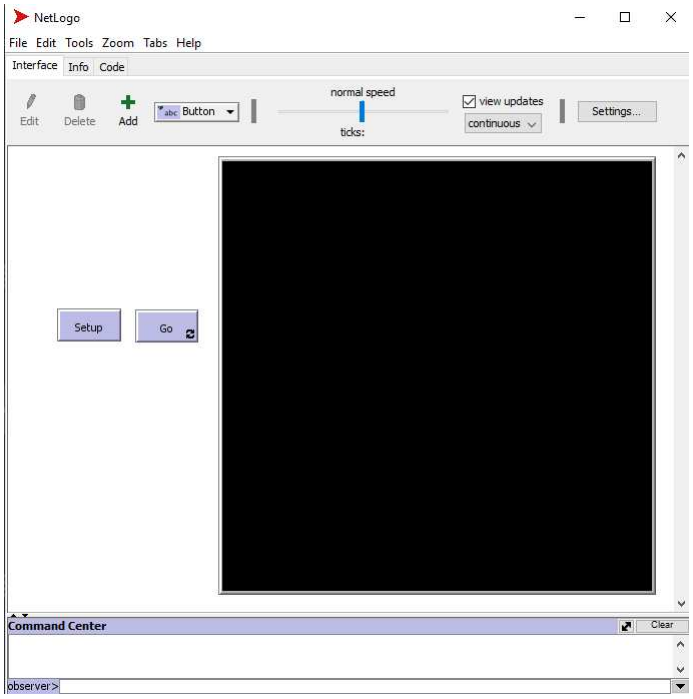


Figure 4.2: Basic Netlogo Environment. Source: (TISUE; WILENSKY, 2004)

As shown in Figure 4.2, most simulations have two buttons on their Interface tab. One is the Setup Button, used to make the model take the initial conditions defined by the code and by the user. The other one is the Run button, which runs the simulation.

4.3 An ABMS for Cervical Cancer Detection

In this research, we built an ABMS for Cervical Cancer. The model used to test the proposed methodology is made to analyze the clinical, social, economic, and behavioral factors that influence women to present cervical cancer in advanced stages, that is, stages in which the carcinoma is extended beyond the cervix (Stages II, III, and IV according to the International Federation of Gynecology and Obstetrics, FIGO). We built this model with the collaboration of the Women's hospital of the University of Campinas (CAISM).

As mentioned in Section 2.1, Cervical cancer is associated to social and economic conditions of countries and regions (BRAY et al., 2012). In Brazil, cancer is the second-largest cause of death (SAÚDE BRASIL, 2019). The expectation for 2020 is 16,4 cervical cancer cases for each 100.000 women in Brazil and 9,6 for each 100.000 women in the state of São Paulo (CÂNCER, 2020).

Cervical cancer is caused mainly by Human Papillomavirus (HPV) (WALBOOMERS et al., 1999). Even though, most HPV infections are quickly cured (SCHIFFMAN; CASTLE; MAUCORT-BOULCH, et al., 2007), persistent infections may increase the risk of precursor lesions that may lead into cervical cancer if not properly treated (SCHIFFMAN; CASTLE; JERONIMO, et al., 2007b).

There are also some clinical co-factors related to cervical cancer, such as the number of sexual partners, smoking, and immunodeficiency (CASTELLSAGUE; BOSCH; MUNOZ, 2002). Another variables, such as fertility and the number of gestation and births, are also associated to a higher risk, even though they are not officially recognized as cofactors (MUÑOZ et al., 2002; LIAO et al., 2012).

A study performed by Vale et al. (2019) showed that states with higher Human Development Index (HDI) , places in which the density of health centers, hospitals, and facilities is higher, more cases are diagnosed at early stage. Cervical cancer can be detected at asymptomatic women at early stage through cervical cancer screening. As public cervical cancer programs increase screening, the proportion of cases at advanced stage tend to fall.

As described earlier, SDoH such as the HDI of the place of residence affect the proportion of cases at advanced stage. Other factors as age (MR, 2016), level of schooling (FRANCESCHI et al., 2009; GYENWALI; PARIYAR; ONTA, 2013), income level, being member of a minority, commuting difficulties or lack of access to a private health plan (POWELL et al., 2018) are also associated to late detection of cervical cancer.

In order to study the incidence of SDoH in the proportion of cases of cervical cancer detected at advanced stage, we designed our ABMS. This ABMS will allow the estimation of the proportion of women diagnosed with cervical cancer at an early stage.

The definition of the variables used in our ABMS depend on the data available. From the datasets, we evaluated the following variables. From the SEADE dataset, having data from all 645 cities in the state of São Paulo, 16 variables were initially considered: city, Municipal HDI (mHDI), Index of Social Responsibility of the State of São Paulo (ISR) for 2014, 2016, and 2018 (ordinal from 1 to 5), fecundity rate for 2010, 2017, and 2018, beds per 1000 people for 2010 and 2017, illiteracy rate, elementary school completion rate, income rate, access to water service rate, access to garbage collection service rate, and access to sewage rate.

For the validation of the model, an ecological study of demographic and socio-economical indicators related to cervical cancer at an advanced stage, from the Hospital Record of the State of São Paulo and from SEADE foundation. This process led to the creation of a logit model, which is used for validation of the ABMS. The resulting logit model has some of these variables as the independent variables and the stage of detection, being 1 the detection at an early stage, and 0, the detection at an advanced stage, as the dependent variable. This logit model was used to determine some variables present in the ABMS and for its validation.

An initial model was built to work as a baseline model. Two entities are present in that model: Women and health facilities of cities. Women can develop symptoms related to cervical cancer and decide whether they go to health facilities. Health facilities, in turn, represent the health system response in the model (hospitals, health centers, etc.). They attend women according to their capacity or to the attention time (working vs non-working hours). Both women and health facilities are located in specific space units (cities). Cities contain specific environmental conditions that determine access to health facilities. In this version, a population from 100 to 3000 women is created into a 8x8 grid. In this version, each city has one health facility.

The state variables for women are shown in Table 4.1. Women's ages range according to Brazilian recommendations for cervical cancer screening. Some clinical co-factors, such as immunodeficiency, sexual initiation age, pregnancy, and number of children are not considered in this project, as we want to keep it as simple as possible. Yet, those factors might be added in future versions.

Table 4.1: State variables for women

Variable name	**Variable type**	**Units**	**Range**	**Meaning**
age	discrete-dynamic	years	25 or more	woman's age
cancer	discrete - dynamic	None	No cancer early advanced	Whether there is cervical cancer and, if positive, its stage.
lastScreening	discrete-dynamic	years	0 or more	Last time the woman was tested
womanLocation	continuous-dynamic	(x; y)	(0; 0) - (7; 7)	Woman's location
Schooling Level	ordinal	None	1-5	Schooling level, being 1 the lowest and 5 the highest

The state variables for health facilities are shown in Table 4.2. In this model, the number of women attended is used as a state variable, as it cannot be calculated from other variables. Also, the health center location is placed at the city's geographical center.

Table 4.2: State variables for health facilities

Variable name	**Variable type**	**Units**	**Range**	**Meaning**
ISR	discrete - static	None	1 - 5	City's ISR
womenAttended	discrete-dynamic	None	0-No limit	Number of women attended at the health center at a given time step
cityLocation	discrete-dynamic	(x; y)	(0; 0) - (7; 7)	City location

In ABMS, time is discrete. In this model, each time step represents one year, as cervical cancer takes several years to grow and screening procedures are recommended each three to five years. In this version of the model, space has no specific units, as it represents an area belonging to each city, and it is not related to physical dimensions of the real world. No dimensions different from space and time are represented in this model.

The environment created for this model is shown in Figure 4.3. Each square of the model represents a city. The lighter the square, the higher its ISR. Also, women are represented as shapes in their exact location.

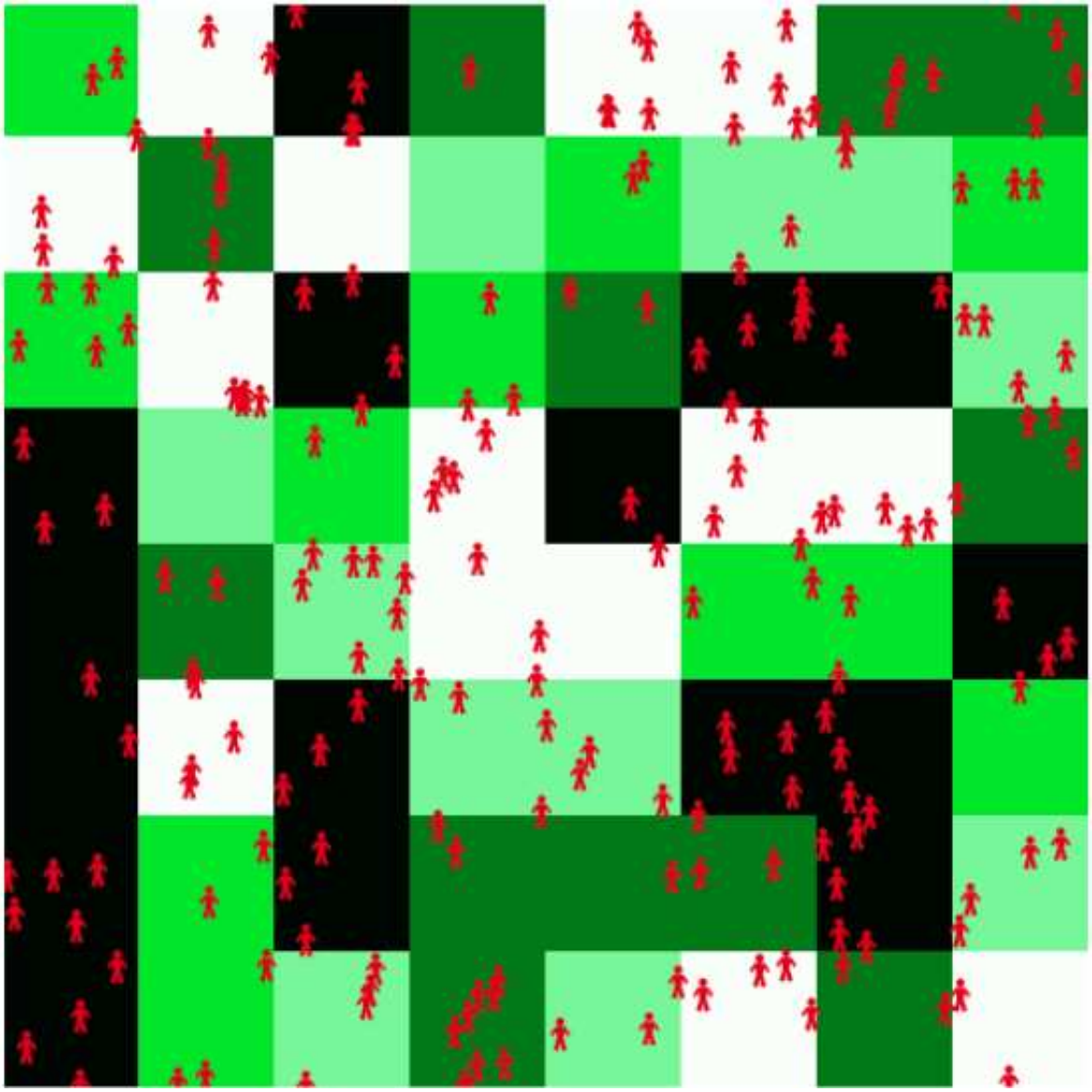


Figure 4.3: Basic Netlogo Environment

Women interact within them in their city by going to the health facility. When one woman decides to go to the health facility, she restricts the number of other women screened

and treated. Health facilities' occupation rate becomes a scarce good women compete for, depending on their intention for having screening. In those mediated interactions, overscreening may be a critical issue. Each health facility interacts with women living in their city by performing screening and treatment.

The objective for facilities consists of detecting cancer at an early stage. Their success is measured by dividing the number of cancer cases detected at an early stage and the total number of cancer cases in the city.

The resulting ABMS is documented following the last version of the ODD protocol (GRIMM; RAILSBACK, et al., 2020). A summary of the ODD protocol for this research project is available at Appendix B.

Chapter 5

Dataset Creation, Preprocessing, and Logistic Regression Results

In this chapter, the dataset creation, the data preprocessing, and the logistic regression results are described.

5.1 Dataset Creation and Data Preprocessing

As mentioned in Chapter 4, two datasets were used to compose a single dataset (FOSP and SEADE datasets). From the SEADE dataset, having data from all 645 cities in the state of São Paulo, 16 variables were initially considered: city, Municipal HDI (mHDI), ISR for 2014, 2016, and 2018 (ordinal from 1 to 5), fecundity rate for 2010, 2017, and 2018, beds per 1000 people for 2010 and 2017, illiteracy rate, elementary school completion rate, income rate, access to water service rate, access to garbage collection service rate, and access to sewage rate.

In order to avoid collinearity issues, and as these variables were highly correlated, for all variables with records for different years, only one year was selected. Considering that cervical cancer stage is developed through time, we chose the oldest record for each case.

From the FOSP dataset, 9502 records of women diagnosed with cervical cancer in the State of São Paulo, from 2010 to 2017, were extracted, to match with the SEADE dataset. It was composed by variables: age, city, schooling level (ordinal from 1 to 5), cervical cancer stage (ordinal from I to IV), and type of lesion (categorical). As data were in different ranges, we performed Z-score normalization (HAN; PEI; TONG, 2022).

Both datasets had missing values. For the SEADE dataset, hospital beds per 1000 people had 287 missing values. For the FOSP dataset, there were two variables with missing data: cervical cancer stage, with 406 values, and schooling level, with 2258 missing values. For the cervical cancer stage, as the number is considerably low, the records with missing data were discarded.

In order to perform the imputation, it is necessary to check if the data are Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR). For data being MCAR, the observed data distribution needs to be similar to missing data distribution. For data being MAR, an observed data variable distribution needs to indicate the missing data distribution. Finally, when data is MNAR, the missing data behavior cannot be defined from observed data and corresponds to a latent phenomenon.

A graphic explanation of the three types of data is shown in Figure 5.1. Even though MCAR and MAR data seem similar, the difference is that in the MCAR case, the relationship between data distributions persists for all variables, whereas in the MAR case, this relationship exists for one or more variables (not all). In the MNAR case, there is no relationship present.

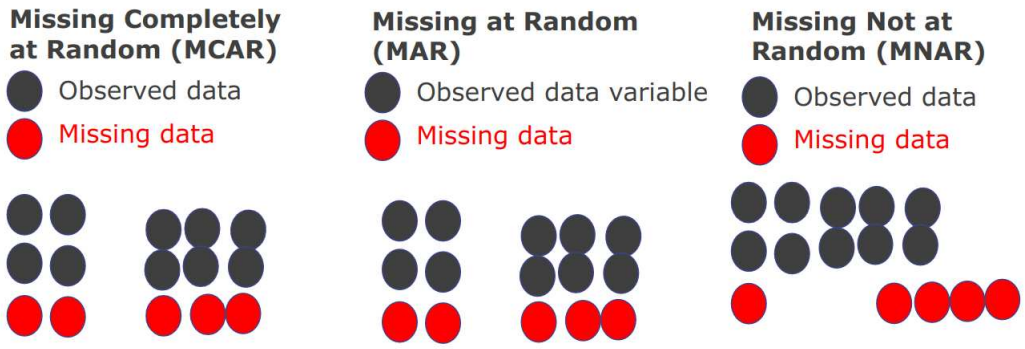


Figure 5.1: Types of Missing Data. Source: Adapted from (GALINDO; FORMIGARI; VALE, D. B., et al., 2021)

For both datasets, we executed the following process. First, we performed a descriptive analysis of missing and observable data to determine if data are MCAR, MAR, or MNAR. We then performed multiple imputation with $m = 20$, using Predictive Mean Matching (PMM), Classification and Regression Trees (CART), and Random Forest (RF) (GALINDO; FORMIGARI; VALE, D. B., et al., 2021).

In order to accept a data imputation algorithm, three parameters, λ, riv, and γ are used.The ratios are calculated considering how large is the amount of the total variance, T, is related

to variance due to imputed data (V), the variance due to sampling (U) and the number of imputations, as shown in Figure 5.2.

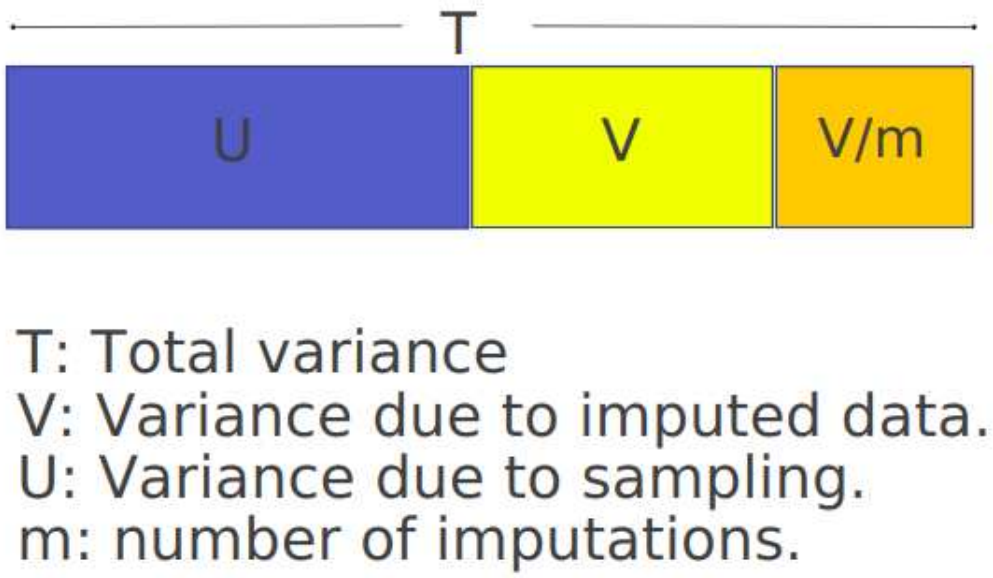


Figure 5.2: Variance ratios for imputation evaluation. Source: Adapted from (GALINDO; FORMIGARI; VALE, D. B., et al., 2021)

The first variance ratio, λ, is the ratio of variance attributable to missing data. It is calculated using equation 5.1:

$$\lambda = \frac{V + V/m}{T} \tag{5.1}$$

The second variance ratio, *riv*, is the relative increase in variance due to nonresponse. It is calculated using equation 5.3:

$$riv = \frac{V + V/m}{U} \tag{5.2}$$

Finally, the third variance ratio, γ, is the proportion of information about the variable to be imputed being missing due to nonresponse. It is calculated using equation 5.3, wirh df being the degrees of freedom present in the data:

$$\gamma = \frac{riv + 2/df + 3}{1 + df} \tag{5.3}$$

These values for all variance ratios are considered modest around 0.2, moderately large around 0.3, and high if above 0.5. In this work, imputations are considered acceptable if they are below 0.2 for all ratios (VAN BUUREN, 2018).

For the FOSP dataset, this process goes as follows. Fig. 5.3 shows the distribution of missing and complete data in Schooling according to age. Missing data distribution appears in red, whereas observed data appears in blue. The lines closer to the edges represent the 75% of the distributions for missing and observed data. According to the relationship between these distributions, there is a clear relationship between missing data and complete data according to patients' age. As the distribution of missing and observable data according to age are similar, it is possible to assume data behavior as MAR.

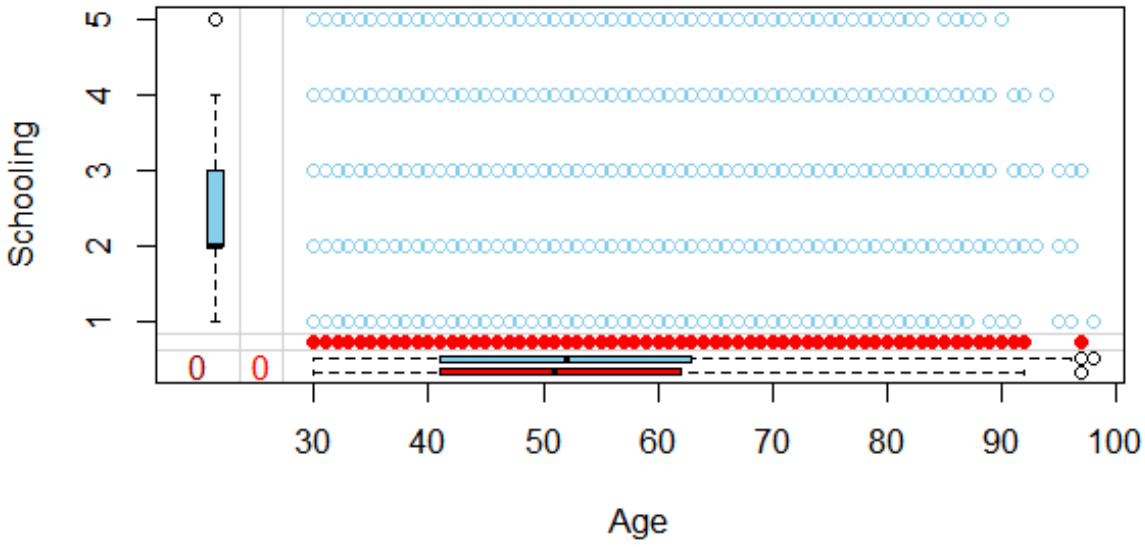


Figure 5.3: Distribution of missing and complete data in patient's schooling level vs patient's age. Source: (GALINDO; FORMIGARI; VALE, D. B., et al., 2021)

Three algorithms were compared for FOSP dataset: PMM, CART, and RF, with $m = 20$ in all cases. Then, a logit model was performed, with the cervical cancer stage as the dependent variable, and schooling level, patient's age, and cancer morphological description as independent variables. We compared these algorithms using the *riv*, λ, and γ variance ratios (Table 5.1). PMM is the only algorithm for which all variance ratios are below 0.2, making it acceptable.

Table 5.1: Data Imputation Comparison for FOSP Dataset

Algorithm	**Variance Ratios**		
	riv	λ	γ
PMM	0.1860334	0.1551585	0.1575775
CART	0.2618193	0.20578635	0.2095631
RF	0.2280557	0.1841833	0.1873557

We followed a similar process for the SEADE dataset. Fig. 5.4 shows the distribution of missing and observed data for hospital beds per 1000 people according to sewage collection rate. It is also possible to see a relationship between missing and observed data, thus allowing the MAR assumption.

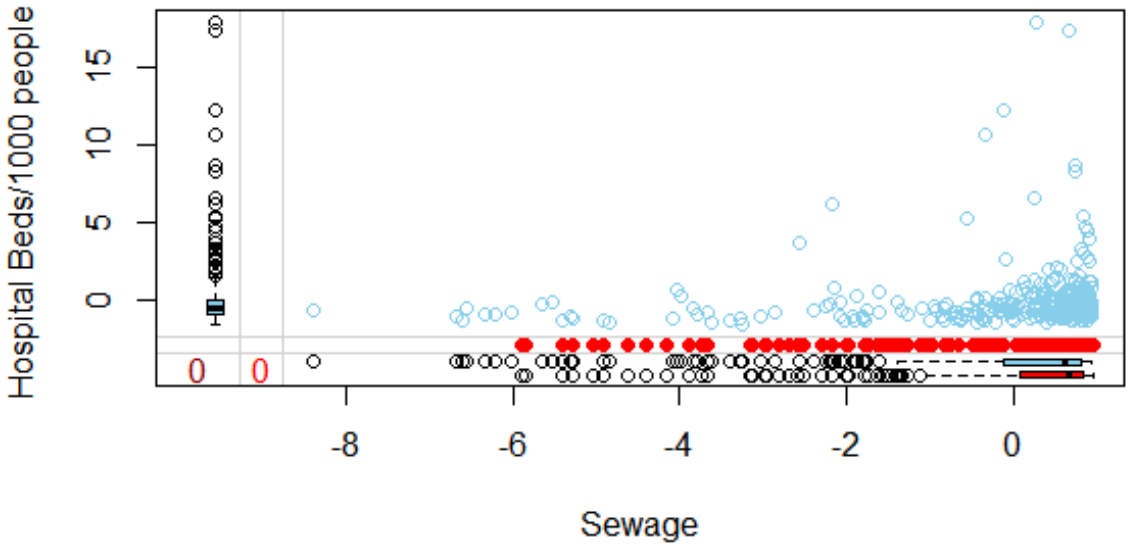


Figure 5.4: Distribution of missing and complete data in hospital beds per 1000 people vs sewage collection rate (z-normal scale). Source: (GALINDO; FORMIGARI; VALE, D. B., et al., 2021)

Four contrasting algorithms were compared: PMM, BayesMI, BootMI, and RF. In all cases, multiple imputation with $m = 20$ was performed and a linear regression with mHDI as the dependent variable was used to compare them. Results for the *riv*, λ, and γ variance ratios appear in Table 5.2. In this case, the only algorithm with acceptable results is BootMI.

Table 5.2: Data Imputation Comparison for FOSP Dataset

Algorithm	**Variance Ratios**		
	riv	λ	γ
PMM	1.174148739	0.540049868	0.556453610
BayesMI	0.988183078	0.497028211	0.512562989
BootMI	1.727693e-05	1.727663e-05	0.005899985
RF	0.748178682	0.427976093	0.441717055

As data were imputed for detecting the impact of SDoH in cervical cancer, both datasets were combined after imputation using PMM for the FOSP dataset and BootMI for the SEADE dataset. A logit regression was performed using beds per 1000 people as the independent variable and the stage of illness as the dependent variable. A p-value of 0,037 was obtained, showing a significant relationship greater than 95%. This is also relevant for the data imputation process itself, as proper data imputation should be useful for general purposes.

5.2 Logistic Regression Modeling

We created two training models. The first model used the Stage as the dependent variable and the second model used the ISR as the dependent variable.

For the first model, 70% of the data was selected for model training and 30% was reserved for testing. Yet, as around 70% of the data corresponds to advanced stage cases, it was necessary to perform some data balancing. Table 5.3 shows the numbers of balanced data.

Table 5.3: Data Balancing

Dataset	**Number of Records**
Total Records	9095
Unbalanced Training data	6366 (1840 Stage I; 4526 Stage II+)
Balanced Training Data	3413 (1722 Stage I; 1691 Stage II+)
Testing Data	2729

To compare the results, we used the following metrics: p-value, chi-square, Akaike Information Criterion (AIC), Sensitivity (TPR), Specificity (TNR), ROC Area under the Curve (ROC-AUC), and finally, Error Rate (ZHENG, 2015). These metrics are defined as follows:

- chi-square(X^2): For a given significance value, X^2 must be higher than a minimum level, depending on the number of degrees of freedom. In this case, for a 0,05 p-value, and 15 degrees of freedom, X^2 must be greater than 25.
- Akaike Information Criterion (AIC): AIC indicates the level of adjustment of the model to the data. Given two models with similar p-values, the lower the AIC the higher the capacity of the model to explain the phenomenon using fewer data, being therefore more appropriate.
- Sensitivity (True Positive Rate - TPR): Capacity of the model to identify correctly positive values of the variable of interest. The closer to 1, the better.
- Specificity (True Negative Rate - TNR): Capacity of the model to identify correctly negative values of the variable of interest. The closer to 1, the better.
- ROC Area under the Curve (AUC): AUC allows summing up the development of a model in terms of the sensitivity and specificity curve, between the values 0.5 and 1. Values closer to 1 indicate higher capacity to identify the values the dependent variable may take.
- Error rate: Classification error rate. The closer to 0, the better.

The comparison between the unbalanced and balanced data is shown in Table 5.4.

Table 5.4: Comparison Between Models Using Unbalanced and Balanced Data

Metric	**Unbalanced Data**	**Balanced Data**
X^2	243.3	219.1
p-value	1.234998e-50	7.674849e-43
AIC	7415	4521.8
AUC	0.6314	0.654
TPR	0.002522068	0.6368222
TNR	1	0.6027893
Error rate	0.2891	0.2833

According to Table 5.4, by balancing data, the value for TPR has a considerable improvement. Also, the overall error rate and AUC are slightly improved. Even though TNR decreased, the objective of this model is to maximize the early detection, making the model

using balanced data more appropriate. Also, a value of TNR equal to 1 and a value of TPR close to 0 show overfitting in the model using unbalanced data. Finally, even though the data balancing affected the other metrics, they are still good metrics to work with. Considering this, and the fact that logit models are widely known and easy to interpret (GUJARATI, 2021), we decided to use the logit model for calibration of the ABMS.

For the balanced model, Table 5.5 shows the individual p-values for all variables, considering a significance level of 5% ($p \leq 0.05$). Variables, such as mHDI did not add significance to the regression and were excluded from the final model. According to the results shown in this table, the ISR was significant for a transition from level 1 to level 2 (ISR_{1-2}) and for level 4 to level 5 (ISR_{4-5}), and obtained a value close to significance for level 3 to level 4(ISR_{3-4}). This result shows the importance of an indicator tuned for the local level as the ISR instead of an indicator used for global comparisons like mHDI. Other significant variables were Elementary School Rate (ElemSchool), and Schooling going above level 2.

Table 5.5: Individual p-values for final logit model

Variable	Age	ISR_{1-2}	ISR_{2-3}	ISR_{3-4}	ISR_{4-5}
p-value	< 2e-16	0.02408	0.09108	0.06383	0.02997
Variable	IncomeRate	Water	Garbage	ElemSchool	Fecundity
p-value	0.93463	0.56033	0.27951	0.02558	0.11402
Variable	$Schooling_{1-2}$	$Schooling_{2-3}$	$Schooling_{3-4}$	$Schooling_{4-5}$	Beds
p-value	0.12184	0.00493	2.06e-06	1.15e-09	0.18648

5.3 Index of Social Responsibility as the dependent variable

Considering the results for the ISR level, we decided to perform an additional logistic regression to test if it is possible to establish the stage of cervical cancer as an indicator of the ISR of the city, and therefore, to consider the proportion of the cervical cancer cases in an early stage as a general indicator of the city's general health scenario. To do this, we performed a logistic regression (GALINDO; FORMIGARI; ZEFERINO, et al., 2023).

For this regression, we also considered the variables, Stage and morphology of the lesion. For analysis, we divided age in two categories: women younger than 50 years old (<50 years

old) and 50 years old or older (>=50 years old). For morphology, we considered three categories: tumors with squamous morphology (SCC) or others (adenocarcinomas, adenosquamous carcinomas, and other types).

The ISR summarizes the situation of each municipality regarding wealth, education and longevity. When combined, these dimensions create a typology that classifies cities into five groups: Dynamic, cities with a high level of wealth and good in social indicators (ISR 5); Unequal, cities that, despite having high levels of wealth, don't achieve good indicators in social dimensions (ISR 4); Equitable, cities with low levels of wealth, but good social indicators (ISR 3); In transition, cities with low levels of wealth and intermediate levels of longevity and/or education (ISR 2); and Vulnerable, the most disadvantaged cities, both in terms of wealth and social indicators (ISR 1).

Table 5.6 shows the distribution of cervical cancer cases among the five ISR groups as a function of the Stage, age and morphology groups, as well as the results of univariate ordinal logistic regressions performed using the ISR as the dependent variable, and the Stage of the disease, the age group (under or over 50), and the tumor morphology, as independent variables in each case. The proportion of cases in Stage 1 (less advanced) increases significantly as the ISR increases, ranging from 24,9% in the group of ISR 1 to 30,0% in the group of cities with ISR 5 (p=0,040). There was no variation in the age group as a function of the ISR, ranging from 54.0% to 60.2% the proportion of cases in women aged 50 years or older (p=0.117). Squamous were the most frequent types of tumors. Yet, their proportion decreased when related to the ISR (p=0.117).

According to the results shown in Table 5.6, the only significant variable for 95% is the stage of cervical cancer. This may indicate that the proportion of cases detected at an early stage could be used as an indicator of the general living conditions of a population in a given city.

To determine how an order increase in ISR may indicate an increase in the chances of detecting cervical cancer in Stage I, we performed a univariate and a multivariate logit models, being the other variables the age-group (divided into lower and higher than 50) and the tumor morphology. Then, we checked at the p-value, the odds-ratio (OR), and the 95% confidence interval (CI). The results are presented in Table 5.7.

Table 5.7 shows that, for each ISR increase, the risk of presenting cervical cancer in Stage I was at least 30% higher than in the immediately previous ISR. In the multivariate analysis,

Table 5.6: Distribution of cervical cancer cases by ISR of the place the women live

ISR	**ISR 1**	**ISR 2**	**ISR 3**	**ISR 4**	**ISR 5**	
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**p-value**
			Total Number of Patients			
	354	839	1101	4137	2664	
			Stage of Cervical Cancer			
Stage 1	88 (24.9)	233 (27.8)	305 (27.7)	1209 (29.2)	798 (30.0)	**0.040**
Stage 2+	266 (75.1)	606 (72.2)	796 (72.3)	2928 (70.8)	1866 (70.0)	
			Age-group			
<50 years	152 (42.9)	334 (39.8)	481 (43.7)	1904 (46.0)	1132 (42.5)	0.117
$\geq$ 50 years	202 (57.1)	505 (60.2)	620 (56.3)	2233 (54.0)	1532 (57.5)	
			Morphology			
SCC	264 (74,6)	628 (74,9)	799 (72,6)	3113 (75,3)	1780 (66,8)	0,117
Others	90 (25,4)	211 (25,1)	302 (27,4)	1024 (24,7)	884 (33,2)	

Table 5.7: Results of Univariate and Multivariate Regressions for Cervical Cancer Stage

			Univariate			**Multivariate**		
Stage								
ISR	**Stage 1**	**Stage 2+**	**p-value**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**
1	88	266	-	-	-	-	-	-
1 vs 2	233	606	0.025	1.36	1.04-1.77	**0.016**	1.40	1.07-1.84
2 vs 3	305	796	0.027	1.33	1.04-1.71	**0.037**	1.31	1.02-1.70
3 vs 4	1209	2928	0.006	1.37	1.10-1.72	**0.011**	1.35	1.07-1.70
4 vs 5	798	1866	0.002	1.43	1.14-1.80	**0.006**	1.39	1.10-1.76

the risk was more significant when comparing ISR 1 and 2, as women living in cities with ISR 2 had a 1.4 times higher risk of being diagnosed in Stage 1 than those living in cities with ISR 1. These results show how the ISR and the Stage of diagnosis are highly related, and then, the ISR can be used as the basis for the ABMS model.

Chapter 6

ABMS Results

In this chapter, the creation and refinement of the ABMS is shown.

6.1 Agent-Based Model for Cervical Cancer

The base Agent-Based Model for cervical cancer follows the process shown in Algorithm 1. This algorithm has two special functions. The first one is cancer_evolution, made to calculate the pass of cancer from stage 0 (no cancer) to stage 1 (early) and from stage 1 to stage 2+ (advanced), depending on the SDoH defined by the logit function modeling and in the time passed since last screening. The second function is cancer_testing, which controls the screening process in each city for all women.

For the first testing of the algorithm, the run cancer evolution was defined as shown in Algorithm 2. It can be seen that in this first version, the chance of cancer evolving from initial to advanced stage, given the patient did not get tested in one year, is 50%. This is an exaggeration, created to test the concept of cancer evolution more easily.

The cancer testing algorithm, for this initial model, was defined as shown in Algorithm 3. In this initial algorithm, as capacity is related to city's ISR, this algorithm makes that women in city's with low ISR may have fewer chances of being diagnosed (GALINDO JARAMILLO et al., 2023).

This algorithm was tested 20000 times. After 20 iterations, for 300 women, the algorithm stabilizes in around 67% of all women with cervical cancer, this cancer being detected at an early stage. The number of women was modified to find the algorithm that best fits the results

Algorithm 1: Cervical Cancer Baseline ABMS

Output: Number of women with cervical cancer at stage 1 and stage 2
Initial Conditions Setting
create 64 cities;
for *each city* **do**
    $ISR \leftarrow randominteger(1, 5)$;
    $capacity \leftarrow 1 + randominteger(1, ISR)$;
create X women;
for *each woman* **do**
    $age \leftarrow randominteger(25, 60)$;
    $cancer \leftarrow 0$;
    location in a city at random;
Simulation Running
while $min - ageof women < 65$ **do**
    increase age of all women by 1;
    set all facilities' womenAttended to 0;
    run cancer_evolution;
    run cancer_testing;
generate-output;

Algorithm 2: Initial Cancer Evolution Algorithm

for *each women* **do**
    SwitchCancer **case** *0* **do**
        $cancer \leftarrow randominteger(0, 1)$;
    case *1* **do**
        if *lastScreening = 0* **then**
            $cancer \leftarrow 1$;
        else
            $cancer \leftarrow randominteger(1, 2)$;
    $cancer \leftarrow randominteger(0, 1)$ **case** *2* **do**
        No Operation

Algorithm 3: Initial Cervical Cancer Testing Algorithm

if $womenAttended \leq capacity$ **then**
    $lastScreening \leftarrow 0$;
    For the city, $womenAttended \leftarrow womenAttended + 1$;
else
    $lastScreening \leftarrow lastScreening + 1$;

in general and for each value of ISR. As shown in Figure 6.1, the proportion change is minimum after 1000 women.

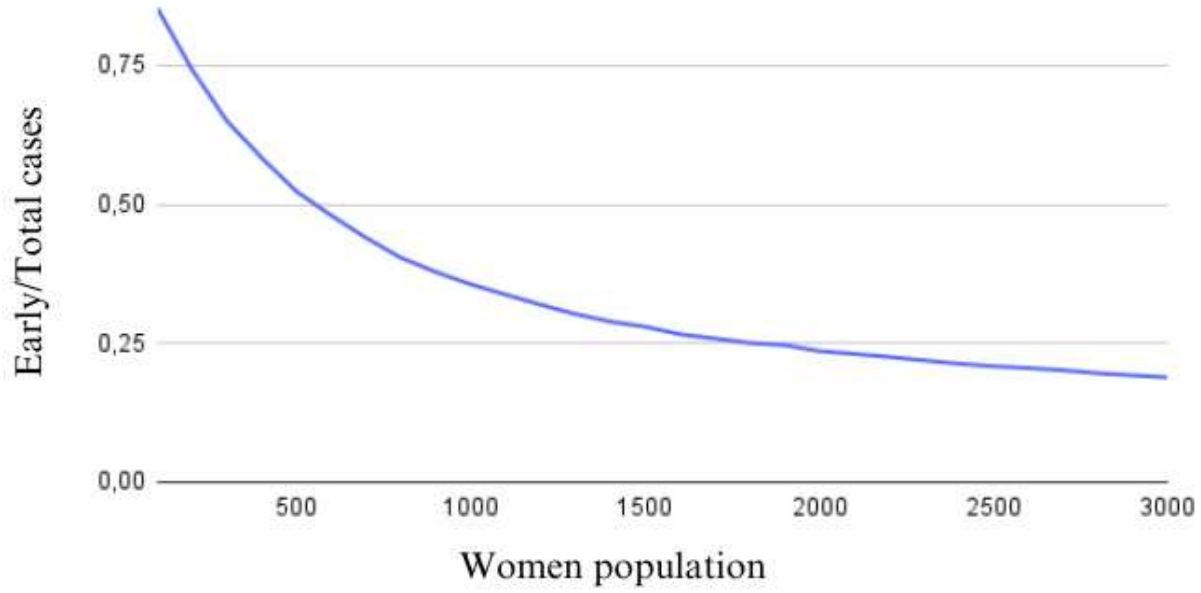


Figure 6.1: Variation of the proportion of Early and Total detected cases according to women population. Source: (GALINDO JARAMILLO et al., 2023)

The results for 1500 women were satisfactory for levels 2 to 5, and also for the total number of cases, as shown in 6.1. These results show that the higher the ISR, the higher the chances of not detecting cervical cancer at an early stage. This represents a macro-level quantitative correspondence between the ABMS and the logit model, being at Level 2 according to Axtell (2005), as there is correspondence with the empirical data for this case.

Stage1/Total	ISR 1	ISR 2	ISR 3	ISR 4	ISR 5	General
300 women	0.380	0.802	0.781	0.667	0.721	0.674
1500 women	0.142	0.205	0.307	0.295	0.316	0.280
Dataset	0.258	0.270	0.290	0.289	0.303	0.289

Table 6.1: Mean Cervical Cancer Cases in Stage 1 divided by Total Cases in for each ISR. Source: (GALINDO JARAMILLO et al., 2023)

As the schooling level was a significant variable, and it may be considered a variable in which municipal efforts may have some effect, we created a random ordinal variable called Schooling, going from 1 to 5, being 1 the lowest level and 5 the highest level. From the dataset, the Pearson correlation level between city's ISR and Schooling is 0.062, instead of considering ISR. Then, we created it considering only the proportion of Schooling level in

population according to the dataset, making women have time to go for screening at working hours, following a study performed in a city of the State of São Paulo, specifically São Jose do Rio Preto, that pointed lack of time as one of the causes for poor participation in cervical cancer screening (IGLESIAS et al., 2019).

To reflect the lack of time to get screened, we created a binary variable for women called working-hours-availability. When working-hours-availability is 1, women can be attended on working hours. When working-hours-availability is 0, in turn, women can only be attended on non-working hours, such as in the evening or in the weekend. At each time step, working-hours-availability is recalculated.

In order to determine the impact of the availability of attention-time for facilities, we created a variable created nonworking-attention-time. Its value can be 0, if there is no nonworking-hours attention time, or 1, if there is. This variable is calculated at the simulation setup, according to the ISR level.

The parameters for working-hours-availability, according to women's schooling level, and nonworking-attention-time were calibrated considering the proportion of early stage cases detected. To do this, we performed 68 rounds of 1021 simulations, varying these two parameters. To select the parameter values, we consider the error rate for each individual stage. The final values for working-hours-availability are shown in Table 6.2. The final nonworking-attention-time values, in turn, are shown in Table 6.3.

Schooling Level	working-hours-availability probability
5	13.0%
4	12.9%
3	12.8%
2	12.5%
1	12.4%

Table 6.2: Women's probability to go for cervical cancer screening in working hours

The results for mean Cervical Cancer Cases in Stage 1 divided by Total Cases for each ISR with the selected parameter values are shown in Table 6.4 For all ISR values except ISR1, and in the general case, the error rate is below 10%, and the higher proportion as the ISR increases, show the model considering Schooling is validated at level 2, according to to Axtell (2005).

These results show the viability of using ABMS for recreating the impact of SDoH within a population. Yet, this work still has some limitations. One of these limitations is the access

ISR Level	Probability of nonworking-attention-time
5	22.0%
4	21.6%
3	21.2%
2	20.8%
1	14.4%

Table 6.3: Attention-time at non-working hours probability according to the ISR level

Stage1/Total	ISR 1	ISR 2	ISR 3	ISR 4	ISR 5	General
Simulations	0.216	0.270	0.276	0.273	0.287	0.272
Logit model	0.258	0.270	0.290	0.289	0,303	0.289
% Err	0.163	0.000	0.048	0.055	0.083	0.059

Table 6.4: Mean Cervical Cancer Cases in Stage 1 divided by Total Cases in for each ISR with selected parameters

to patient data related to working-hours-availability and to cities' attention hours, necessary to achieve level three validation. These values may affect the results of the overall model.

One thing to notice about this model is that small variations in probability of being available for testing on working hours for women and in attention-time at non-working hours probability for facilities lead to changes in the proportion of cervical cancer cases detected at an early stage. This is due to the complexity of the model and lead space for improvement in early detection through few changes in the long run. This makes this model fit to analyze through Reinforcement Learning for facilities.

As a final form of validation, and in order to see how our model behaves in presence of further parameter variations, we run one factor at a time (OFAT) sensitivity analysis, including two parameters: women-population, going from 500 to 3000 in intervals of 50 women, and nonworking-attention-facilities, going from 0 to 64 in intervals of 1. The default values for women-population and for nonworking-attention-facilities are 1500 and 12, respectively.

The results for the distribution of the proportion between Early and total detected cases for women population analysis are shown in Figure 6.2. The mean value and the variance for the proportion of Early/Total detected cases are 0.263 and

$$1.68 * 10^{-5}$$

respectively. These results show that the model has little variation regarding the number of women.

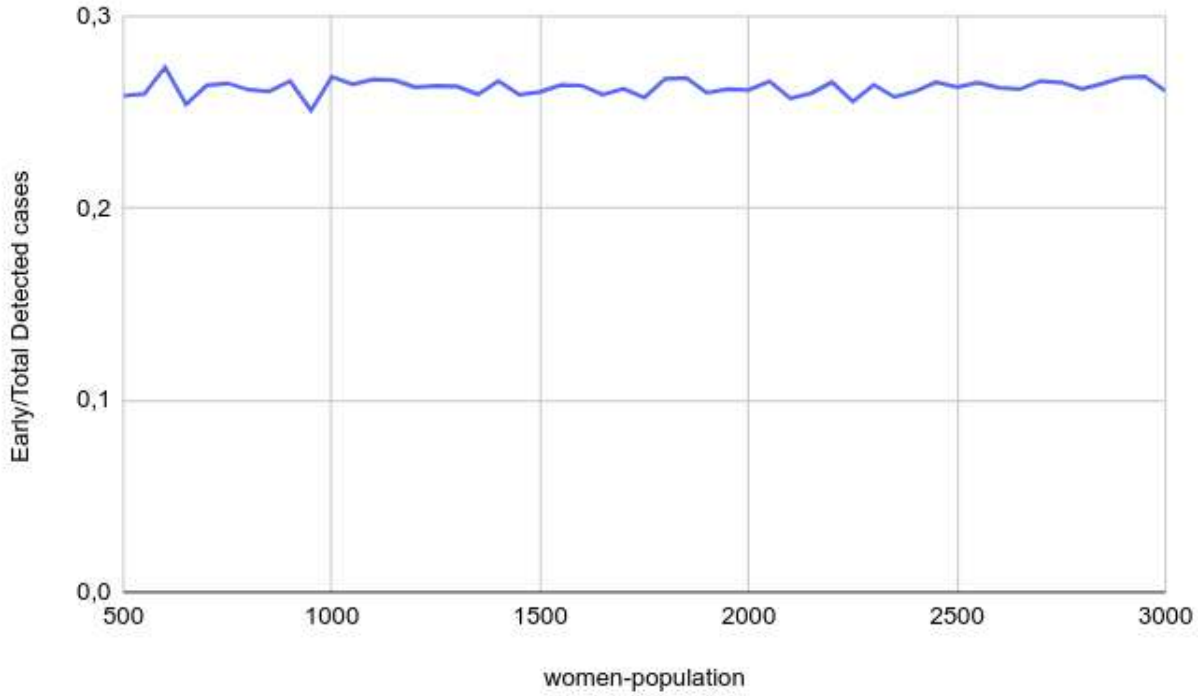


Figure 6.2: Variation of the proportion of Early and Total detected cases according to women population

The results for the distribution of the proportion between Early and total detected cases for the number of facilities with nonworking-hours attention are shown in Figure 6.3. The mean value and the variance for the proportion of Early/Total detected cases are 0.583 and

$$6.88 * 10^{-2}$$

respectively. It is possible to see that the results are highly affected by the availability of non-working hours attention. For example, the proportion goes from 0.299 at 14 facilities with non-working attention to 0.360 at 15 facilities. These abrupt changes show that this parameter is prone to increase women's early detection. Then, changes in this parameter show promising results for machine learning approaches, such as Reinforcement Learning. Yet, because of the high variance resulting on increasing this value for the overall model, these changes may affect the model's stability.

These results show that our model is able to quantify the incidence of SDoH in the detection of cervical cancer at an early stage. This result might be useful to point out possible policies

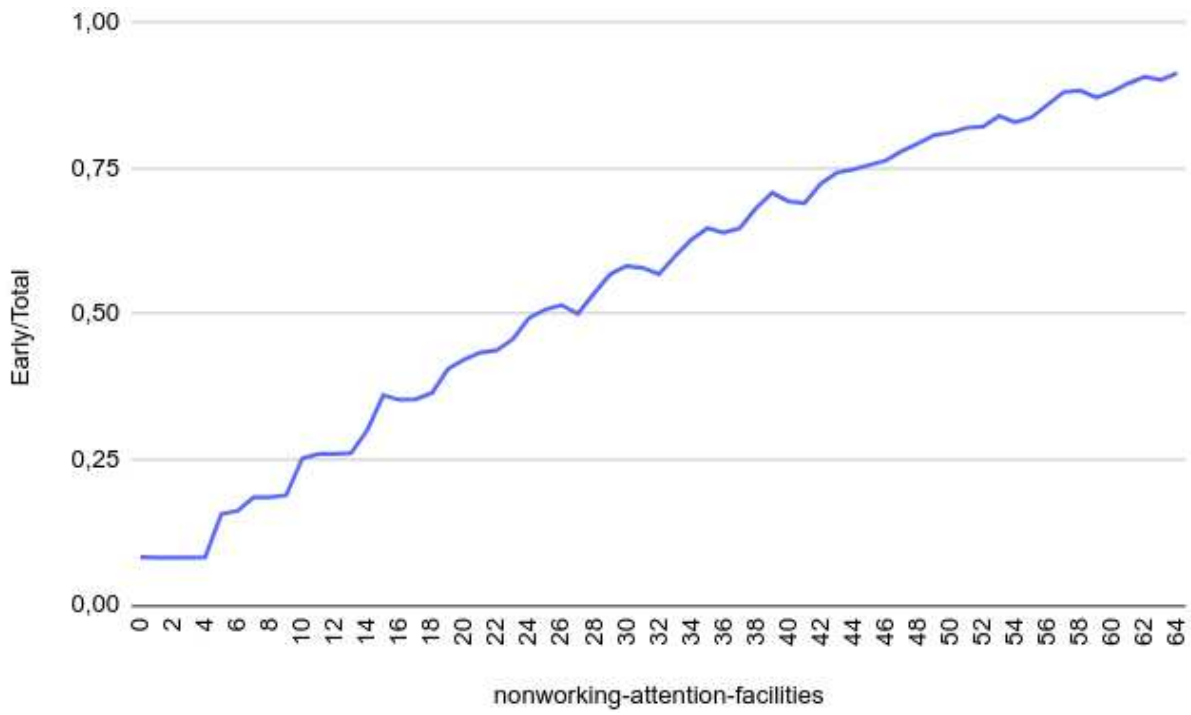


Figure 6.3: Variation of the proportion of Early and Total detected cases according to the number of nonworking facilities

to maximize early stage detection. Yet, testing and validating these policies is limited by the lack of access to data on the receptiveness of patients to different measures. With these data, the use of machine learning techniques to recreate patient behavior may be possible, allowing the testing of several scenarios.

Chapter 7

Conclusions

The general objective of this research was to develop a method to systematically model and evaluate public health scenarios related to cervical cancer detection at an early stage, considering SDoH and using ABMS. The results obtained show that a method consisting in creating a reference model, create a baseline ABMS, and validate and expand the ABMS using the reference model is a valid framework for early detection of cervical cancer within a population.

One of the main contributions of this work was the quantification of the incidence of SDoH in the detection of cervical cancer at an early stage. The significance of ISR in the logit model shows the importance of environmental and social conditions for prevention and early detection of diseases. Further understanding of the incidence of SDoH in cervical cancer and other diseases is necessary for better public policy definition, monitoring, and execution.

Another contribution is the ABMS itself. This model allows the quantification of the impact of the actions taken by decision-makers at a city level and the possible responses from patients considering the social conditions in which they are inserted. Further refining of this ABMS may be useful for practitioners in decision-making processes for public health policy.

There are some key considerations for each step on this method. For the use of the reference model, we made some choices that helped to achieve this goal. First, we created the reference model using real and trustable data. This is a key advantage, yet this might not be possible for all cases. Second, we decided to use a classic, simple, and easy-to-explain method, the logistic regression. There are more complex models that may result in better performance. Considering the scope of this project, we preferred explainability over performance. Yet, we obtained significant results using a logit model.

For the creation of the baseline ABMS, we decided to keep our model as simple as possible. By doing this, we were able to understand the responses of agents to different scenarios (occupation and schedule availability). For the last stage, model validation and calibration, having few parameters helped to understand their incidence in our model. Yet, for future use of the model, it may be necessary to add more variables and parameters.

There were some limitations for our model. The main limitation is intrinsic to interdisciplinary work, and it is related to the fact that in model creation, it is necessary to make assumptions. Even though our ABMS shows internal consistency and the use of the logit model gave external validation, it needs further testing with real data, especially with policymakers insights and data from patient behaviors. This may be necessary for the implementation of policies using our model.

Another limitation is the time it takes to make changes and test different scenarios in the current version of our ABMS. Future versions of the model may include artificial intelligence algorithms to define agent behavior. Yet, the inclusion of these techniques may affect model performance, as the number of agents with these algorithms increase.

Finally, future works include the collection of patient data and policymaker insights to expand this ABMS. This may imply the creation of more reference models for validation and calibration. The use of artificial intelligence, especially reinforcement learning techniques, sees promising to recreate agent behaviors. As the inclusion of more variables, more reference models and artificial intelligence techniques may affect model performance, future model architecture may include parallelization. Finally, our methodology may be adapted to other diseases, for which public datasets are available.

Bibliographic References

ALVAREZ-GALVEZ, J.; SUAREZ-LLEDO, V. Using Agent-Based Modeling to Understand the Emergence and Reproduction of Social Inequalities in Health. In: 1. MULTIDISCIPLINARY Digital Publishing Institute Proceedings. [S.l.: s.n.], 2019. v. 44, p. 2.

ARTHUR, W. B. Inductive reasoning and bounded rationality. **The American economic review**, JSTOR, v. 84, n. 2, p. 406–411, 1994.

AXTELL, R. **Three distinct kinds of empirically-relevant agent-based models**. [S.l.], 2005.

AXTELL, R.; AXELROD, R.; EPSTEIN, J. M.; COHEN, M. D. Aligning simulation models: A case study and results. **Computational & mathematical organization theory**, Springer, v. 1, n. 2, p. 123–141, 1996.

BADHAM, J.; BARBROOK-JOHNSON, P.; CAIADO, C.; CASTELLANI, B. Justified Stories with Agent-Based Modelling for Local COVID-19 Planning. **Journal of Artificial Societies and Social Simulation**, v. 24, n. 1, p. 8, 2021. ISSN 1460-7425. DOI: 10.18564/jasss.4532. Available from: <http://jasss.soc.surrey.ac.uk/24/1/8.html>.

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999.

BRAY, F.; JEMAL, A.; GREY, N.; FERLAY, J.; FORMAN, D. Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. **The lancet oncology**, Elsevier, v. 13, n. 8, p. 790–801, 2012.

BUSS, P. M.; PELLEGRINI FILHO, A. A saúde e seus determinantes sociais. **Physis: revista de saúde coletiva**, SciELO Public Health, v. 17, p. 77–93, 2007.

CÂNCER, N. .-. I. N. de. **Conceito e Magnitude**. [S.l.: s.n.], Aug. 2020. Available from: <https://www.inca.gov.br/controle-do-cancer-do-colo-do-utero/conceito-e-magnitude>.

CAREY, G.; CRAMMOND, B. Systems change for the social determinants of health. **BMC public health**, BioMed Central, v. 15, n. 1, p. 1–10, 2015.

CASTELLSAGUE, X.; BOSCH, F. X.; MUNOZ, N. Environmental co-factors in HPV carcinogenesis. **Virus research**, Elsevier, v. 89, n. 2, p. 191–199, 2002.

CHIN, K. O.; GAN, K. S.; ALFRED, R.; ANTHONY, P.; LUKOSE, D. Agent architecture: An overview. **Transactions on science and technology**, v. 1, n. 1, p. 18–35, 2014.

CONTE, R.; PAOLUCCI, M. On agent-based modeling and computational social science. **Frontiers in Psychology**, Frontiers, v. 5, p. 668, 2014.

COOLEY, P.; SOLANO, E. Agent-based model (ABM) validation considerations. In: PROCEEDINGS of the Third International Conference on Advances in System Simulation (SIMUL 2011). [S.l.: s.n.], 2011. P. 134–139.

DAVID, N.; SICHMAN, J.; COELHO, H.; MARIETTO, M. The structure and logic of interdisciplinary research in agent-based social simulation. **Journal of Artificial Societies and Social Simulation**, University of Surrey, n. 3, 2004.

DENNY, L.; SANJOSE, S. de; MUTEBI, M.; ANDERSON, B. O.; KIM, J.; JERONIMO, J.; HERRERO, R.; YEATES, K.; GINSBURG, O.; SANKARANARAYANAN, R. Interventions to close the divide for women with breast and cervical cancer between low-income and middle-income countries and high-income countries. en. **Lancet**, England, v. 389, n. 10071, p. 861–870, Nov. 2016.

DURÁN, J. M. The Universe of Computer Simulations. In: COMPUTER Simulations in Science and Engineering. [S.l.]: Springer, 2018. P. 1–29.

EPSTEIN, J. M.; AXTELL, R. **Growing artificial societies: social science from the bottom up**. [S.l.]: Brookings Institution Press, 1996.

FERLAY, J.; SOERJOMATARAM, I.; ERVIK, M.; DIKSHIT, R.; ESER, S.; MATHERS, C.; REBELO, M.; PARKIN, D.; FORMAN, D.; BRAY, F. International agency for research on cancer. In: GLOBOCAN 2012 v10, cancer incidence and mortality worldwide: IARC CancerBase no 11 globocaniarcfr. [S.l.: s.n.], 2013.

FRANCESCHI, S.; PLUMMER, M.; CLIFFORD, G.; DE SANJOSE, S.; BOSCH, X.; HERRERO, R.; MUNOZ, N.; VACCARELLA, S. Differences in the risk of cervical cancer and human papillomavirus infection by education level. **British journal of cancer**, Nature Publishing Group, v. 101, n. 5, p. 865–870, 2009.

FUENTES, M. Methods and methodologies of complex systems. **Modeling complex systems for public policies**, IPEA Brasılia, DF, p. 55–72, 2015.

FURTADO, B. A.; SAKOWSKI, P. A. M.; TÓVOLLI, M. H. A complexity approach for public policies. Instituto de Pesquisa Econômica Aplicada (Ipea), 2015.

GALINDO, J. F.; FORMIGARI, G. M.; VALE, D. B.; URSINI, E. L.; MARTINS, P. S. Missing Data: Comparison of Multiple-Imputation Algorithms for Social Determinants of Health in Cervical Cancer Stage Detection. In: 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Vancouver, BC, Canada: IEEE, Oct. 2021. P. 0509–0514. ISBN 978-1-66540-066-4. DOI: 10.1109/IEMCON53756.2021.9623097. Available from: <https://ieeexplore.ieee.org/document/9623097/>. Visited on: 17 Nov. 2022.

GALINDO, J. F.; FORMIGARI, G. M.; ZEFERINO, L. C.; CARVALHO, C. F.; URSINI, E. L.; VALE, D. B. Social determinants influencing cervical cancer diagnosis: an ecological study. **International Journal for Equity in Health**, BioMed Central, v. 22, n. 1, p. 1–7, 2023.

GALINDO JARAMILLO, J. F.; GRANDO, L.; LEITE, J. R. E.; VALE, D. B.; URSINI, E. L. In: INFORMS. 2023 Winter Simulation Conference Proceedings. [S.l.: s.n.], 2023.

GAO, J.; BARZEL, B.; BARABÁSI, A.-L. Universal resilience patterns in complex networks. **Nature**, Nature Publishing Group, v. 530, n. 7590, p. 307, 2016.

GENTILE, J.; GLAZNER, C.; KOEHLER, M. Simulation models for public policy. **Modeling Complex Systems For Public Policies**, p. 73–83, 2015.

GENTILE, J. E.; DAVIS, G. J.; RUND, S. S. Verifying agent-based models with steady-state analysis. **Computational and Mathematical Organization Theory**, Springer, v. 18, n. 4, p. 404–418, 2012.

GIABBANELLI, P. J.; CRUTZEN, R. Using agent-based models to develop public policy about food behaviours: future directions and recommendations. **Computational and mathematical methods in medicine**, Hindawi, v. 2017, 2017.

GRIMM, V.; BERGER, U.; BASTIANSEN, F.; ELIASSEN, S.; GINOT, V.; GISKE, J.; GOSS-CUSTARD, J.; GRAND, T.; HEINZ, S. K.; HUSE, G., et al. A standard protocol for describing individual-based and agent-based models. **Ecological modelling**, Elsevier, v. 198, n. 1-2, p. 115–126, 2006.

GRIMM, V.; BERGER, U.; DEANGELIS, D. L.; POLHILL, J. G.; GISKE, J.; RAILSBACK, S. F. The ODD protocol: a review and first update. **Ecological modelling**, Elsevier, v. 221, n. 23, p. 2760–2768, 2010.

GRIMM, V.; POLHILL, G.; TOUZA, J. Documenting Social Simulation Models: The ODD Protocol as a Standard. In: **Simulating Social Complexity: A Handbook**. Ed. by Bruce Edmonds and Ruth Meyer. Cham: Springer International Publishing, 2017. P. 349–365. ISBN 978-3-319-66948-9. DOI: 10.1007/978-3-319-66948-9_15. Available from: <https://doi.org/10.1007/978-3-319-66948-9_15>.

GRIMM, V.; RAILSBACK, S. F.; VINCENOT, C. E.; BERGER, U.; GALLAGHER, C.; DEANGELIS, D. L.; EDMONDS, B.; GE, J.; GISKE, J.; GROENEVELD, J., et al. The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. **Journal of Artificial Societies and Social Simulation**, v. 23, n. 2, 2020.

GUJARATI, D. N. **Essentials of econometrics**. [S.l.]: Sage Publications, 2021.

GYENWALI, D.; PARIYAR, J.; ONTA, S. R. Factors associated with late diagnosis of cervical cancer in Nepal. **Asian Pac J Cancer Prev**, v. 14, n. 7, p. 4373–7, 2013.

HAN, J.; PEI, J.; TONG, H. **Data mining: concepts and techniques**. [S.l.]: Morgan kaufmann, 2022.

HOGAN, J. W.; GALAI, N.; DAVIS, W. W. Modeling the impact of social determinants of health on HIV. **AIDS and Behavior**, Springer, v. 25, p. 215–224, 2021.

IGLESIAS, G. A.; LARRUBIA, L. G.; NETO, A. d. S. C.; PACCA, F. C.; IEMBO, T. Conhecimento e adesão ao Papanicolau de mulheres de uma rede de Atenção Primária à Saúde. **Revista de Ciências Médicas**, v. 28, n. 1, p. 21–30, 2019.

INCA. **Detecção Precoce**. [S.l.: s.n.], Sept. 2022. Available from: <https://www.gov.br/inca/pt-br/assuntos/gestor-e-profissional-de-saude/controle-do-cancer-do-colo-do-utero/acoes/deteccao-precoce>.

INCA. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero / Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede**. 2. ed. [S.l.]: Ministério da Saúde, Instituto Nacional do Câncer (INCA), 2016. ISBN 978-85-7318-296-5.

INCA. **Ficha técnica de indicadores das ações de controle do câncer do colo do útero**. [S.l.]: INCA Rio de Janeiro, 2014.

KHADEMI, A.; ZHANG, D.; GIABBANELLI, P. J.; TIMMONS, S.; LUO, C.; SHI, L. An agent-based model of healthy eating with applications to hypertension. In: ADVANCED Data Analytics in Health. [S.l.]: Springer, 2018. P. 43–58.

KIESLING, E.; GÜNTHER, M.; STUMMER, C.; WAKOLBINGER, L. M. Agent-based simulation of innovation diffusion: a review. **Central European Journal of Operations Research**, Springer, v. 20, n. 2, p. 183–230, 2012.

KLÜGL, F.; BAZZAN, A. L. Agent-based modeling and simulation. **AI Magazine**, v. 33, n. 3, p. 29–29, 2012.

KOEHLER, M.; BARRY, P.; MEYER, T. Sending agents to war. In: PROCEEDINGS of the Agent 2006 Conference. Chicago, IL, Argonne National Lab. [S.l.: s.n.], 2006.

LANGELLIER, B. A.; LÊ-SCHERBAN, F.; PURTLE, J. Funding quality pre-kindergarten slots with Philadelphia's new 'sugary drink tax': simulating effects of using an excise tax to address a social determinant of health. **Public Health Nutrition**, Cambridge University Press, v. 20, n. 13, p. 2450–2458, 2017.

LIAO, S.-F.; LEE, W.-C.; CHEN, H.-C.; CHUANG, L.-C.; PAN, M.-H.; CHEN, C.-J. Baseline human papillomavirus infection, high vaginal parity, and their interaction on cervical cancer risks after a follow-up of more than 10 years. **Cancer Causes & Control**, Springer, v. 23, n. 5, p. 703–708, 2012.

LINEWEAVER, C. H.; DAVIES, P. C.; RUSE, M. What is complexity? Is it increasing. **Complexity and the Arrow of Time**, Cambridge University Press, p. 3–15, 2013.

MITCHELL, M. **Complexity: A guided tour**. [S.l.]: Oxford University Press, 2011.

MORSHED, A. B.; KASMAN, M.; HEUBERGER, B.; HAMMOND, R. A.; HOVMAND, P. S. A systematic review of system dynamics and agent-based obesity models: Evaluating obesity as part of the global syndemic. **Obesity Reviews**, Wiley Online Library, v. 20, p. 161–178, 2019.

MR, N. C. R. P. I. of. **Three year report of the Population Based Cancer Registries 2012–14.** [S.l.: s.n.], Aug. 2016. Available from: <http://www.ncrpindia.org/WOMEN/Annual_Reports.aspx>.

MUÑOZ, N.; FRANCESCHI, S.; BOSETTI, C.; MORENO, V.; HERRERO, R.; SMITH, J. S.; SHAH, K. V.; MEIJER, C. J.; BOSCH, F. X.; RESEARCH ON CANCER (IARC) MULTICENTRIC

CERVICAL CANCER STUDY GROUP, I. A. for, et al. Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. **The Lancet**, Elsevier, v. 359, n. 9312, p. 1093–1101, 2002.

OH, H.; TRINH, M. P.; VANG, C.; BECERRA, D. Addressing Barriers to Primary Care Access for Latinos in the US: An Agent-Based Model. **Journal of the Society for Social Work and Research**, The University of Chicago Press Chicago, IL, v. 11, n. 2, p. 000–000, 2020.

PAGE, S. E. **Diversity and Complexity**. 1. ed. [S.l.]: Princeton University Press, 2010. (Primers in Complex Systems). ISBN 0691137676,9780691137674.

POWELL, T. C.; DILLEY, S. E.; BAE, S.; STRAUGHN JR, J. M.; KIM, K. H.; LEATH III, C. A. The Impact of Racial, Geographic and Socioeconomic Risk Factors on the Development of Advanced Stage Cervical Cancer. **Journal of lower genital tract disease**, NIH Public Access, v. 22, n. 4, p. 269, 2018.

RAILSBACK, S. F.; GRIMM, V. **Agent-based and individual-based modeling: a practical introduction**. [S.l.]: Princeton university press, 2019.

RAND, W. Complex systems: concepts, literature, possibilities and limitations. **Modeling complex systems for public policies. Brasılia: IPEA**, p. 37–54, 2015.

RAND, W.; BROWN, D. G.; PAGE, S. E.; RIOLO, R.; FERNANDEZ, L. E.; ZELLNER, M., et al. Statistical validation of spatial patterns in agent-based models. In: PROCEEDINGS of agent based simulation. [S.l.: s.n.], 2003. v. 4.

RASELLA, D.; MORAIS, G. A. d. S.; ANDERLE, R. V.; SILVA, A. F. d.; LUA, I.; COELHO, R.; RUBIO, F. A.; MAGNO, L.; MACHADO, D.; PESCARINI, J., et al. Evaluating the impact of social determinants, conditional cash transfers and primary health care on HIV/AIDS: Study protocol of a retrospective and forecasting approach based on the data integration with a cohort of 100 million Brazilians. **Plos one**, Public Library of Science San Francisco, CA USA, v. 17, n. 3, e0265253, 2022.

SALIB, M. Y.; RUSSELL, J. H. B.; STEWART, V. R.; SUDDERUDDIN, S. A.; BARWICK, T. D.; ROCKALL, A. G.; BHARWANI, N. 2018 FIGO Staging Classification for Cervical Cancer: Added Benefits of Imaging. en. **RadioGraphics**, v. 40, n. 6, p. 1807–1822, Oct. 2020. ISSN 0271-5333, 1527-1323. DOI: 10.1148/rg.2020200013. Available from: <http://pubs.rsna.org/doi/10.1148/rg.2020200013>. Visited on: 20 Dec. 2022.

SAÚDE BRASIL, M. da. **TabNet Win32 3.0: Mortalidade - Brasil.** [S.l.]: Ministério da Saúde Brasil, Dec. 2019. Available from: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt10uf.def>.

SCHELLING, T. C. Models of segregation. **The American Economic Review**, JSTOR, v. 59, n. 2, p. 488–493, 1969.

SCHIFFMAN, M.; CASTLE, P. E.; JERONIMO, J.; RODRIGUEZ, A. C.; WACHOLDER, S. Human papillomavirus and cervical cancer. **The Lancet**, v. 370, n. 9590, p. 890–907, 2007. ISSN 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(07)61416-0. Available from: <https://www.sciencedirect.com/science/article/pii/S0140673607614160>.

SCHIFFMAN, M.; CASTLE, P. E.; JERONIMO, J.; RODRIGUEZ, A. C.; WACHOLDER, S. Human papillomavirus and cervical cancer. **The Lancet**, Elsevier, v. 370, n. 9590, p. 890–907, 2007.

SCHIFFMAN, M.; CASTLE, P. E.; MAUCORT-BOULCH, D.; WHEELER, C. M.; UNDETERMINED SIGNIFICANCE/LOW-GRADE SQUAMOUS INTRAEPITHELIAL LESIONS TRIAGE STUDY) GROUP, A. (S. C. of; PLUMMER, M. A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. **The Journal of infectious diseases**, The University of Chicago Press, v. 195, n. 11, p. 1582–1589, 2007.

SEADE. **Indice Paulista de Responsabilidade Social -IPRS METODOLOGIA**. [S.l.: s.n.], Feb. 2016. P. 28. Available from: <https://iprs.seade.gov.br/downloads/pdf/metodologia_do_iprs_2018.pdf>. Visited on: 24 Feb. 2023.

SICHMAN, J. S. Operationalizing Complex Systems. **Modeling Complex Systems For Public Policies**, p. 85–123, 2015.

SILVA, G. A.; ALCANTARA, L. L. d. M.; TOMAZELLI, J. G.; RIBEIRO, C. M.; GIRIANELLI, V. R.; SANTOS, É. C.; CLARO, I. B.; ALMEIDA, P. F. d.; LIMA, L. D. d. Avaliação das ações de controle do câncer de colo do útero no Brasil e regiões a partir dos dados registrados no Sistema Único de Saúde. **Cadernos de Saúde Pública**, SciELO Public Health, v. 38, e00041722, 2022.

STARR, J.; KAIN, M. Agent-Based Simulation of Social Determinants of Health for Equitable COVID-19 Intervention. In: 2022 7th International Conference on Intelligent Informatics and Biomedical Science (ICIIBMS). [S.l.: s.n.], 2022. v. 7, p. 319–326. DOI: 10.1109/ICIIBMS55689.2022.9971638.

STONEDAHL, F.; WILENSKY, U. Finding forms of flocking: Evolutionary search in abm parameter-spaces. In: SPRINGER. MULTI-AGENT-BASED Simulation XI: International Workshop, MABS 2010, Toronto, Canada, May 11, 2010, Revised Selected Papers 11. [S.l.: s.n.], 2011. P. 61–75.

TEMKIN, S. M.; RIMEL, B.; BRUEGL, A. S.; GUNDERSON, C. C.; BEAVIS, A. L.; DOLL, K. M. A contemporary framework of health equity applied to gynecologic cancer care: A Society of Gynecologic Oncology evidenced-based review. en. **Gynecologic Oncology**, v. 149, n. 1, p. 70–77, Apr. 2018. ISSN 00908258. DOI: 10.1016/j.ygyno.2017.11.013. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0090825817315081>. Visited on: 17 Nov. 2022.

TEN BROEKE, G.; VAN VOORN, G.; LIGTENBERG, A. Which sensitivity analysis method should I use for my agent-based model? **Journal of Artificial Societies and Social Simulation**, v. 19, n. 1, p. 5, 2016.

TISUE, S.; WILENSKY, U. Netlogo: A simple environment for modeling complexity. In: BOSTON, MA. INTERNATIONAL conference on complex systems. [S.l.: s.n.], 2004. v. 21, p. 16–21.

TRACY, M.; CERDÁ, M.; KEYES, K. M. Agent-based modeling in public health: current applications and future directions. **Annual review of public health**, Annual Reviews, v. 39, p. 77–94, 2018.

VALE, D.; SAUVAGET, C.; MURILLO, R.; MUWONGE, R.; ZEFERINO, L.; SANKARANARAYANAN, R. Correlation of Cervical Cancer Mortality with Fertility, Access to Health Care and Socioeconomic Indicators. en. **Revista Brasileira de Ginecologia e Obstetrícia / RBGO Gynecology and Obstetrics**, v. 41, n. 04, p. 249–255, Apr. 2019. ISSN 0100-7203, 1806-9339. DOI: 10.1055/s-0039-1683859. Available from: <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0039-1683859>. Visited on: 17 Nov. 2022.

VALE, D. B.; TEIXEIRA, J. C.; BRAGANÇA, J. F.; DERCHAIN, S.; SARIAN, L. O.; ZEFERINO, L. C. Elimination of cervical cancer in low- and middle-income countries: Inequality of access and fragile healthcare systems. en. **International Journal of Gynecology & Obstetrics**, v. 152, n. 1, p. 7–11, Jan. 2021. ISSN 0020-7292, 1879-3479. DOI: 10.1002/ijgo.13458. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ijgo.13458>. Visited on: 9 Jan. 2023.

VALE, D. B.; SAUVAGET, C.; MUWONGE, R.; THULER, L. C. S.; BASU, P.; ZEFERINO, L. C.; SANKARANARAYANAN, R. Level of human development is associated with cervical cancer stage at diagnosis. **Journal of Obstetrics and Gynaecology**, Taylor & Francis, v. 39, n. 1, p. 86–90, 2019.

VAN BUUREN, S. **Flexible imputation of missing data**. [S.l.]: CRC press, 2018.

VERMEER, W. H.; SMITH, J. D.; WILENSKY, U.; BROWN, C. H. High-fidelity agent-based modeling to support prevention decision-making: An open science approach. **Prevention Science**, Springer, p. 1–12, 2022.

WALBOOMERS, J. M.; JACOBS, M. V.; MANOS, M. M.; BOSCH, F. X.; KUMMER, J. A.; SHAH, K. V.; SNIJDERS, P. J.; PETO, J.; MEIJER, C. J.; MUÑOZ, N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. **The Journal of pathology**, Wiley Online Library, v. 189, n. 1, p. 12–19, 1999.

WEIMER, C. W.; MILLER, J. O.; HILL, R. R. Agent-based modeling: an introduction and primer. In: IEEE. 2016 Winter Simulation Conference (WSC). [S.l.: s.n.], 2016. P. 65–79.

WHITEHEAD, M. The concepts and principles of equity and health. **Health promotion international**, Oxford University Press, v. 6, n. 3, p. 217–228, 1991.

WILENSKY, U. NetLogo flocking model. **Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL**, 1998.

WILENSKY, U.; RAND, W. **An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo**. [S.l.]: MIT Press, 2015.

WORLD HEALTH ORGANIZATION. **Cervical cancer**. en. [S.l.: s.n.], Feb. 2022. publisher: World Health Organization. Available from: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer>. Visited on: 11 Nov. 2022.

ZHENG, A. **Evaluating machine learning models: a beginner's guide to key concepts and pitfalls**. [S.l.]: O'Reilly Media, 2015.

Appendix A

First Page of the Ethics Committee Approval

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: FATORES RELACIONADOS COM O DIAGNÓSTICO TARDIO DO CÂNCER DO COLO DO ÚTERO

Pesquisador: Diama Bhadra Andrade Peixoto do Vale

Área Temática:

Versão: 2

CAAE: 42657020.1.0000.5404

Instituição Proponente: Hospital da Mulher Prof. Dr. José Aristodemo Pinotti - CAISM

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.597.863

Apresentação do Projeto:

As informações contidas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram obtidas dos documentos apresentados para apreciação ética e das informações inseridas pelo Pesquisador Responsável do estudo na Plataforma Brasil.

Introdução: O câncer do colo de útero é uma importante causa de morte entre as mulheres em todo o mundo, e sua ocorrência está associada às condições de desenvolvimento dos países e suas regiões1. No Brasil, desde o ano 2000, o câncer é a segunda maior causa de morte, atrás apenas das doenças cardiovasculares2. Para 2020 são esperados cerca de 16,4 casos a cada 100.000 mulheres no Brasil, e 9,6 casos a cada 100.000 mulheres no estado de São Paulo3.O desenvolvimento do câncer do colo do útero é resultado de infecções persistentes de tipos oncogênicos do papiloma vírus humano (HPV)4, ainda que a maioria das infecções por HPV sejam curadas rapidamente5. A infecção persistente de um dos aproximadamente 15 genótipos oncogênicos aumenta significativamente o risco de lesões precursoras, que se não tratadas podem evoluir para o câncer invasivo6. Cofatores clínicos importantes são alto número de parceiros sexuais, início da vida sexual em idades mais jovens, tabagismo e infecção por HIV ou outra imunodeficiência7. A fertilidade e/ou número de gestações e partos também foram associados a um maior risco. Entretanto nenhum desses é reconhecido como fator independente para a progressão mais acelerada do câncer8–10. Um estudo que reuniu 51.158 casos de câncer de colo

Appendix B

ODD Protocol for the First ABMS

This document describes the Cervical Cancer in society ABM according to the ODD protocol (Grimm et al., 2020).

B.1 Purpose and patterns

he purpose of this specific model is to test the impact of socio-economic factors in cervical cancer in advanced stages, that is, stages in which the carcinoma is extended beyond the cervix (Stages II, III, and IV according to the International Federation of Gynecology and Obstetrics, FIGO). As cervical cancer can be easily treated when detected at Stage I, the existence of a relationship between Social Determinants of Health (SDoH) and the stage of detection of cervical cancer is proposed as an indicator of wellbeing within the population. In this version of the model, the Index of Social Responsibility (ISR) of the State of São Paulo, Brazil, is used as an independent variable. Also, the age and the fertility rate of the patient are used as control variables. Then, the criterion for evaluating the model's suitability for its purpose is the existence or not of a direct relationship between the ISR of the city of residence of women and the proportion of cases in advanced stages. The model will be assumed useful if it can reproduce the existence of that relationship. If that pattern is not reproduced, it may indicate that some important variables are missing or inadequately represented, indicating a necessity of model expansion.

B.2 Entities, state variables, and scales

The state variables for women are shown in Table B.1. Women's ages range according to Brazilian recommendations for cervical cancer screening. Some clinical co-factors, such as immunodeficiency, sexual initiation age, pregnancy, and number of children are not considered in this version, as we want to keep it as simple as possible. Yet, those factors might be added in future versions.

Table B.1: State variables for women

Variable name	Variable type	Units	Range	Meaning
age	discrete - dynamic	years	25-65	Woman's age.
cancer	discrete - dynamic	–	0 - No cancer; 1 - Early; 2 - Advanced	Whether there is cervical cancer and if that cancer is at an early or advanced stage.
lastScreening	discrete - dynamic	years	0 - no limit	Last time the patient was screened
womanLocation	discrete - dynamic	(x; y)	(0; 0) - (7; 7)	Woman's location.

The state variables for cities are shown in Table 2. In this model, the number of women attended is used as a state variable, as it cannot be calculated from other variables. Cities' occupancy level and capacity are also used in this model, but they are calculated from other variables, as shown in Elements 3 and 5 of this document, respectively. For cities location, each square of the grid in which the simulation happens represents a different city. Each city has a specific ISR that goes from 1 to 5, the higher, the better. In this model, the number of women attended is used as a state variable, as it cannot be calculated from other variables

Table B.2: State variables for health facilities

Variable name	Variable type	Units	Range	Meaning
ISR	discrete - static	None	1 - 5	City's ISR
womenAttended	discrete-dynamic	None	0-No limit	Number of women attended at the health center at a given time step
cityLocation	discrete-dynamic	(x; y)	(0; 0) - (7; 7)	City location

Scales: In this model, time is discrete. Each time step represents one year, as cervical cancer takes several years to grow, and screening procedures are recommended each three to five years. In this version of the model, space has no specific units, as it is not related to physical dimensions of the real world. No dimensions different from space and time are represented in this model.

B.3 Process overview and scheduling

The main process for this model goes as follows: Each time step, women's variables age and lastScreening are increased by one year, and cities' womenAttended and occupancy level are set to 0. After that, for each city from (0; 0) to (7; 7), each woman runs the following actions in a random order:

a. Determine if the woman has cervical cancer and, if positive, its stage. In this model, disease remission is not considered. If her prior cancer level is 0, a random number between 0 and 1 is set. If her prior cancer level is 1 and

b. If occupancy level is below 100%, the woman is tested through the following steps.

   (a) LastScreening is set to 0.

   (b) Cities' womenAttended is increased by 1.

After each woman makes the actions listed above, the occupancy level for the corresponding health center is calculated by using equation (1): occupancyLevel = womenAttended/capacity (1)

In this version of the model, the main process consists of determining the impact of determinants of health in the occurrence of cervical cancer in advanced stages. Specifically, the impact of ISR in that occurrence. Even though ISR is not explicitly present in the process described above, it determines the capacity of the cities to attend women, as shown in model initialization (Element 5). That capacity also affects the occupancy level, which, in turn, determines if screening occurs. The order for women to perform the actions described in the process is random and determined each time step to avoid order for being a determinant for the occurrence of advanced stage cancer. Also, a woman can only access the health facility located in her city to avoid distortion on the impact of ISR into advanced stage detection.

B.4 Design concepts

is model is made to incorporate clinical factors and Social Determinants of Health into cervical cancer modeling. This is the first version of the model, making necessary to use ground-based evidence for further testing. Regarding SDoH, this model is made to identify how the ISR of the place of residence of women and the proportion of cases in advanced stages. There is evidence showing that regions with low ISR have a higher proportion of cases of cervical cancer in advanced stages. That may be explained by lower density of facilities for diagnosis and treatment in areas with lower ISR and by the influence of human development in human behavior (Vale et al. 2018). As this is the first model for cervical cancer, of the density of facilities, it is focused on testing the incidence. The incidence of HDI in the density of facilities is considered by defining facilities' capacity as a function of the HDI for each city.

It is expected to emerge an inverse relationship between a city's ISR and the proportion between cases in advanced stages and all cases in that city. Another expected result (acting as a validation of the model) is a direct relationship between women's age and the presence of cancer.

Cities decide whether they receive a woman for cervical screening depending on their capacity. Health facilities objectives are to diagnose and women without exceeding capacity (Element 3); Capacity is decided as a simplification of the relationship between the density of health facilities and the ISR of the city (Element 5).

For women, the fitness function consists of the number of years without cervical cancer in late state. For each woman, of the variable HealthyTime is made by the following

condition: if cancer = no, then increase HealthyTime by 1 Regarding cities, the fitness function consists of being as efficient as possible, attending the most women they can, as fast as they can, and attending all women requiring the service. The rationale behind this subprocess is health facilities need to attend all women they need, avoiding overscreening, that is, screening women more times than necessary and effectively screening all women in their area (Vale et al. 2019b). Yet, the final outcome is to minimize the proportion of women's cancer detected at stage 2.

Learning is not considered in this model as it is the base for studies of impact of SDoH in cervical cancer, and then we intend to keep it as simple as possible. Yet, it is built in order to implement learning techniques for both women and health facilities in future models.

Cities can sense their own occupation rate and the presence of HPV and cancer for all women are attended. Bounded rationality appears as women cannot sense facilities' occupation rate and health facilities cannot sense if a specific woman has cancer to prioritize her. As this model shows the reality of Brazilian health system, health facilities also cannot sense women's last screening.

Women interact within them in their city by going to the health facility. When one woman decides to go to the health facility, she restricts the number of other women screened and treated. Health facilities' occupation rate becomes a scarce good women compete for. Each health facility interacts with women living in their city by performing screening and treatment.

Stochasticity is present in the following processes during model initialization: (a) women's distribution over the environment and age, and also cities SRI are completely random; (b) each health facility's capacity is based on a random number depending on city's SRI, as it is expected that cities with higher SRI have a greater capacity in their health facilities, but there is no direct relationship between those two variables, and; (c) cancer level is also defined randomly, as explained in Element 3.

To test the impact of the ISR of the city in the proportion of cases of cancer in advanced stages, a histogram considering those variables is the main output. The proportion of cases will also be shown graphically by using a color scale, in which the darker the tone in a specific city, the higher the proportion of cases, allowing for the inspection of specific places. For those outputs, it is necessary to collect the number of cases of cancer in early and advanced stages for each city.

B.5 Initialization

At the beginning of the simulation, the space is divided into a 8x8 grid, being each square of the grid a city. The static variables are assigned to each city, recreating the scenario in the state of São Paulo, according to SEADE data. Each city also holds the capacity of attending a determined quantity of women using a logit function with a maximum number equal to the city's ISR.

Women are created and distributed in the environment randomly. Their variables are initialized according to Table B.3. No variations related to cancer are considered to reduce the dependence on initial conditions during simulation.

Variable name	Value
age	Random (25,40)
cancer	0
lastScreening	0
womanLocation	(random; random)

Table B.3: Initial conditions for women

B.6 Input data

This version of the model does not use input data to represent time-varying processes. Yet, it uses the results of a logistic regression using data from São Paulo Oncocentro Foundation and the SEADE Foundation.

B.7 Submodels

This model does not consider submodels, as it is intended as a base model.

Appendix C

Publications Related to this Thesis

As a consequence of this thesis, we published one journal paper, one conference full paper, three conference extended abstracts, and one conference short abstract. As a means to get access to these works, in this Appendix, I will give some information about the paper, and show its first page as an image.

C.1 Missing Data: Comparison of Multiple-Imputation Algorithms for Social Determinants of Health in Cervical Cancer Stage Detection

Type: Conference pap **Year:** 2021 **Authors:** Juan Fernando Galindo Jaramillo, Giovana Moura Formigari; Diama Bhadra Vale; Edson L. Ursini; Paulo Martins. **Conference:** IEMCON 2021. **Available at:** https://ieeexplore.ieee.org/document/9623097

C.2 Fatores Relacionados com o Diagnóstico Tardio do Câncer de Colo de Útero

Type: Extended Abstract. **Year:** 2021 **Authors:** Giovana Moura Formigari; Juan Fernando Galindo Jaramillo; Edson L. Ursini; Diama Bhadra Vale. **Conference:** XXIX Congresso de Iniciação Científica da UNICAMP. **Available at:** https://www.prp.unicamp.br/inscricao-congresso/resumos/2021P18106A35826O5329.pdf

Missing Data: Comparison of Multiple-Imputation Algorithms for Social Determinants of Health in Cervical Cancer Stage Detection

Juan F. Galindo Jaramillo
IEEE member
School of Technology
University of Campinas
Limeira, Brazil
orcid.org/0000-0003-2406-5100

Giovana Moura Formigari
Gynecology and Obstetrics Department
University of Campinas
Campinas, Brazil
g216817@dac.unicamp.br

Diama Bhadra Vale
Gynecology and Obstetrics Department
University of Campinas)
Campinas, Brazil
orcid.org/0000-0003-2423-022

Edson L. Ursini
IEEE member
School of Technology
University of Campinas
Limeira, Brazil
http://orcid.org/0000-0002-1597-4057

Paulo S. Martins
School of Technology
University of Campinas
Limeira, Brazil
orcid.org/0000-0002-8303-5564

***Abstract*—Social Determinants of Health impact general health conditions within a population. However, missing data affect statistical analysis and forecasting of diseases. Multiple imputation has gained momentum and several machine learning algorithms have been used for data imputation. As most statistical analysis and machine learning software have already implemented these algorithms, their performance is usually taken for granted without further analysis. Furthermore, we notice a discrepancy between how imputation must be carried out and how it is usually performed in real-word practice. Thus, in this work we examine different machine learning algorithms for multiple imputation in two datasets with Social Determinants of Health in Cervical Cancer. The results of this comparison are presented.**

***Index Terms*—SDoH, Multiple Imputation, MICE, Imputed Data, Cervical Cancer.**

I. Introduction

The Social Determinants of Health (SDoH) are social and economic factors involved in how individuals work and live [1]. SDoH influence all steps of illness and health. The relationship between SDoH and diseases is relevant to the understanding of those diseases and to suggest strategies for their control. SDoH indicators are usually those generated by governmental agencies of information. Commonly some of these indicators lack data completeness, especially in low- and middle-income countries [1]. To overcome this issue is necessary to define and evaluate proper missing data processes.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Cervical cancer is the fourth most common female cancer in the world and Brazil. Four in every five women in Brazil are diagnosed in advanced stages [2]. SDoH are essential factors influencing access to diagnosis, and its analysis may help to overcome this scenario. Unfortunately, missing data has been a barrier to produce knowledge in the field.

Missing data can be classified into missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) data [3]. Missing data are MCAR when data distribution is similar to the observed data. MAR data, in turn, occurs when the distribution of missing data is related to some property of the observed data. Finally, when missing data behavior cannot be defined, it is said data are MNAR. MCAR data are easier to handle, yet it is rare to find the MCAR case when working with real data.

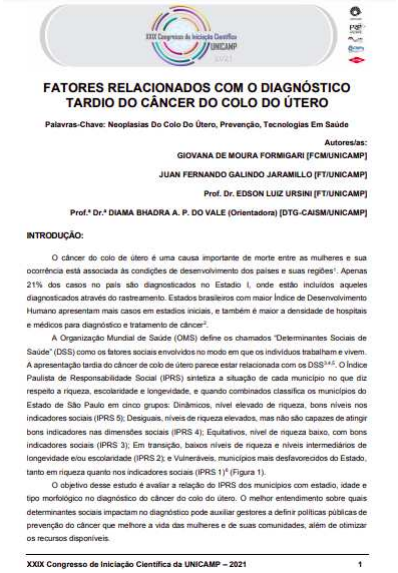
Even though almost all popular data imputation techniques work with MAR data, it is necessary to check how these techniques respond to each case [4]. In general, data imputation is accomplished without considering the behavior of the missing data.

Ad-hoc data imputation techniques, such as imputing data using the mean value or using a regression to impute missing values, are easily found in practice. Yet, such approaches implicitly assume that the behavior of missing data is completely known, thus leading to biased imputations [5].

Proper missing data imputation requires techniques that consider data diversity as the true values of missing data will always be unknown, such as multiple imputation [6]. Multiple imputation techniques simulate different scenarios. To to this, multiple (m) datasets with imputed (simulated) data are created [7].

978-1-6654-0066-4/21/$31.00 ©2021 IEEE

Figure C.1: Missing Data



FATORES RELACIONADOS COM O DIAGNÓSTICO TARDIO DO CÂNCER DO COLO DO ÚTERO

Palavras-Chave: Neoplasias Do Colo Do Útero, Prevenção, Tecnologias Em Saúde

Autores/as:
GIOVANA DE MOURA FORMIGARI [FCM/UNICAMP]
JUAN FERNANDO GALINDO JARAMILLO [FT/UNICAMP]
Prof. Dr. EDSON LUIZ URSINI [FT/UNICAMP]
Prof.ª Dr.ª DIAMA BHADRA A. P. DO VALE (Orientadora) [DTG-CAISM/UNICAMP]

INTRODUÇÃO:

O câncer do colo de útero é uma causa importante de morte entre as mulheres e sua ocorrência está associada às condições de desenvolvimento dos países e suas regiões[1]. Apenas 21% dos casos no país são diagnosticados no Estadio I, onde estão incluídos aqueles diagnosticados através do rastreamento. Estados brasileiros com maior Índice de Desenvolvimento Humano apresentam mais casos em estadios iniciais, e também é maior a densidade de hospitais e médicos para diagnóstico e tratamento de câncer[2].

A Organização Mundial de Saúde (OMS) define os chamados "Determinantes Sociais de Saúde" (DSS) como os fatores sociais envolvidos no modo em que os indivíduos trabalham e vivem. A apresentação tardia do câncer de colo de útero parece estar relacionada com os DSS[3,4,5]. O Índice Paulista de Responsabilidade Social (IPRS) sintetiza a situação de cada município no que diz respeito a riqueza, escolaridade e longevidade, e quando combinados classifica os municípios do Estado de São Paulo em cinco grupos: Dinâmicos, nível elevado de riqueza, bons níveis nos indicadores sociais (IPRS 5); Desiguais, níveis de riqueza elevados, mas não são capazes de atingir bons indicadores nas dimensões sociais (IPRS 4); Equitativos, nível de riqueza baixo, com bons indicadores sociais (IPRS 3); Em transição, baixos níveis de riqueza e níveis intermediários de longevidade e/ou escolaridade (IPRS 2); e Vulneráveis, municípios mais desfavorecidos do Estado, tanto em riqueza quanto nos indicadores sociais (IPRS 1)[6] (Figura 1).

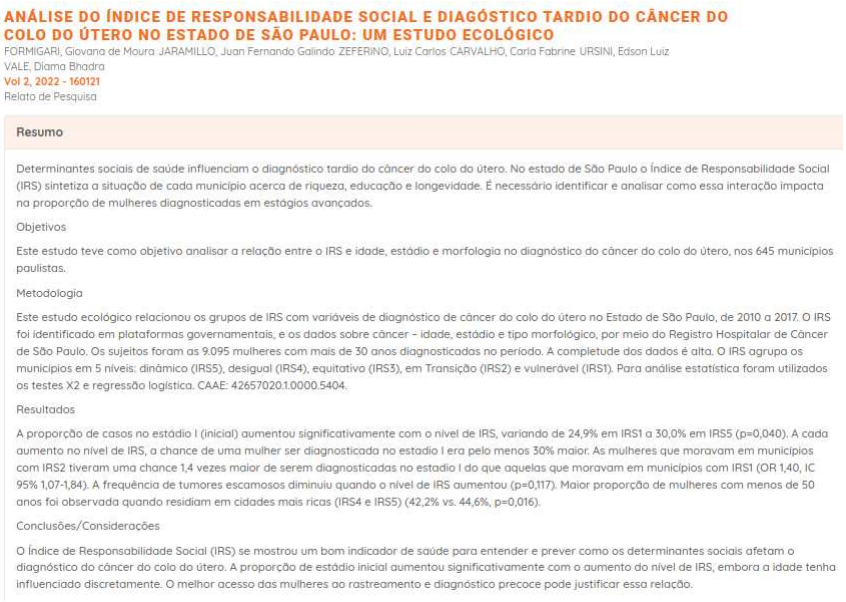
O objetivo desse estudo é avaliar a relação do IPRS dos municípios com estadio, idade e tipo morfológico no diagnóstico do câncer do colo do útero. O melhor entendimento sobre quais determinantes sociais impactam no diagnóstico pode auxiliar gestores a definir políticas públicas de prevenção do câncer que melhore a vida das mulheres e de suas comunidades, além de otimizar os recursos disponíveis.

Figure C.2: Cervical Cancer Related Factors Paper

C.3 ANÁLISE DO ÍNDICE DE RESPONSABILIDADE SOCIAL E DIAGÓSTICO TARDIO DO CÂNCER DO COLO DO ÚTERO NO ESTADO DE SÃO PAULO: UM ESTUDO ECOLÓGICO

Type: Abstract. **Year:** 2022 **Authors:** Giovana Moura Formigari; Juan Fernando Galindo Jaramillo; Luis Carlos Zeferino; Carla Fabrine Carvalho; Edson L. Ursini; Diama Bhadra Vale.

Conference: 13º Congresso Brasileiro de Saúde Coletiva **Available at:** https://proceedings.science/abrascao-2022/trabalhos/analise-do-indice-de-responsabilidade-social-e-diagostico-tardio-do-cancer-do-co?lang=pt-br



ANÁLISE DO ÍNDICE DE RESPONSABILIDADE SOCIAL E DIAGÓSTICO TARDIO DO CÂNCER DO COLO DO ÚTERO NO ESTADO DE SÃO PAULO: UM ESTUDO ECOLÓGICO

FORMIGARI, Giovana de Moura JARAMILLO, Juan Fernando Galindo ZEFERINO, Luiz Carlos CARVALHO, Carla Fabrine URSINI, Edson Luiz VALE, Diama Bhadra

Vol 2, 2022 - 160121

Relato de Pesquisa

Resumo

Determinantes sociais de saúde influenciam o diagnóstico tardio do câncer do colo do útero. No estado de São Paulo o Índice de Responsabilidade Social (IRS) sintetiza a situação de cada município acerca de riqueza, educação e longevidade. É necessário identificar e analisar como essa interação impacta na proporção de mulheres diagnosticadas em estágios avançados.

Objetivos

Este estudo teve como objetivo analisar a relação entre o IRS e idade, estádio e morfologia no diagnóstico do câncer do colo do útero, nos 645 municípios paulistas.

Metodologia

Este estudo ecológico relacionou os grupos de IRS com variáveis de diagnóstico de câncer do colo do útero no Estado de São Paulo, de 2010 a 2017. O IRS foi identificado em plataformas governamentais, e os dados sobre câncer – idade, estádio e tipo morfológico, por meio do Registro Hospitalar de Câncer de São Paulo. Os sujeitos foram as 9.095 mulheres com mais de 30 anos diagnosticadas no período. A completude dos dados é alta. O IRS agrupa os municípios em 5 níveis: dinâmico (IRS5), desigual (IRS4), equitativo (IRS3), em Transição (IRS2) e vulnerável (IRS1). Para análise estatística foram utilizados os testes X2 e regressão logística. CAAE: 42657020.1.0000.5404.

Resultados

A proporção de casos no estádio I (inicial) aumentou significativamente com o nível de IRS, variando de 24,9% em IRS1 a 30,0% em IRS5 (p=0,040). A cada aumento no nível de IRS, a chance de uma mulher ser diagnosticada no estadio I era pelo menos 30% maior. As mulheres que moravam em municípios com IRS2 tiveram uma chance 1,4 vezes maior de serem diagnosticadas no estadio I do que aquelas que moravam em municípios com IRS1 (OR 1,40, IC 95% 1,07-1,84). A frequência de tumores escamosos diminuiu quando o nível de IRS aumentou (p=0,117). Maior proporção de mulheres com menos de 50 anos foi observada quando residiam em cidades mais ricas (IRS4 e IRS5) (42,2% vs. 44,6%, p=0,016).

Conclusões/Considerações

O Índice de Responsabilidade Social (IRS) se mostrou um bom indicador de saúde para entender e prever como os determinantes sociais afetam o diagnóstico do câncer do colo do útero. A proporção de estádio inicial aumentou significativamente com o aumento do nível de IRS, embora a idade tenha influenciado discretamente. O melhor acesso das mulheres ao rastreamento e diagnóstico precoce pode justificar essa relação.

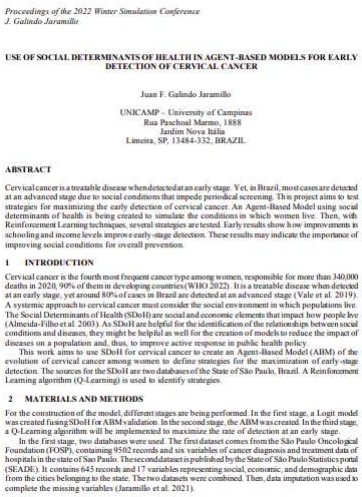
Figure C.3: Index of Social Responsibility Abstract

C.4 Use of Social Determinants of Health in Agent-based Models for Early Detection of Cervical Cancer

Type: Extended Abstract. **Year:** 2022 **Authors:** Juan Fernando Galindo Jaramillo **Conference:** Winter Simulation Conference 2022. **Available at:** https://informs-sim.org/wsc22papers/doc115.pdf

C.5 Social determinants influencing cervical cancer diagnosis: an ecological study

Type: Paper published in Journal. **Year:** 2023 **Authors:** Juan Fernando Galindo Jaramillo, Giovana Moura Formigari; Luiz Carlos Zeferino; Carla Fabrine Carvalho; Edson Luiz Ursini; Diama Bhadra Vale **Journal:** International Journal for Equity in Health **Available at:** https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-023-01912-8



Proceedings of the 2022 Winter Simulation Conference
J. Galindo Jaramillo

USE OF SOCIAL DETERMINANTS OF HEALTH IN AGENT-BASED MODELS FOR EARLY DETECTION OF CERVICAL CANCER

Juan F. Galindo Jaramillo

UNICAMP - University of Campinas
Rua Paschoal Marmo, 1888
Jardim Nova Itália
Limeira, SP, 13484-332, BRAZIL

ABSTRACT

Cervical cancer is a treatable disease when detected at an early stage. Yet, in Brazil, most cases are detected at an advanced stage due to social conditions that impede periodical screening. This project aims to test strategies for maximizing the early detection of cervical cancer. An Agent-Based Model using social determinants of health is being created to simulate the conditions in which women live. Then, with Reinforcement Learning techniques, several strategies are tested. Early results show how improvements in schooling and income levels improve early-stage detection. These results may indicate the importance of improving social conditions for overall prevention.

1 INTRODUCTION

Cervical cancer is the fourth most frequent cancer type among women, responsible for more than 340,000 deaths in 2020, 90% of them in developing countries(WHO 2022). It is a treatable disease when detected at an early stage, yet around 80% of cases in Brazil are detected at an advanced stage (Vale et al. 2019). A systemic approach to cervical cancer must consider the social environment in which populations live. The Social Determinants of Health (SDoH) are social and economic elements that impact how people live (Almeida-Filho et al. 2003). As SDoH are helpful for the identification of the relationships between social conditions and diseases, they might be helpful as well for the creation of models to reduce the impact of diseases on a population and, thus, to improve active response in public health policy.

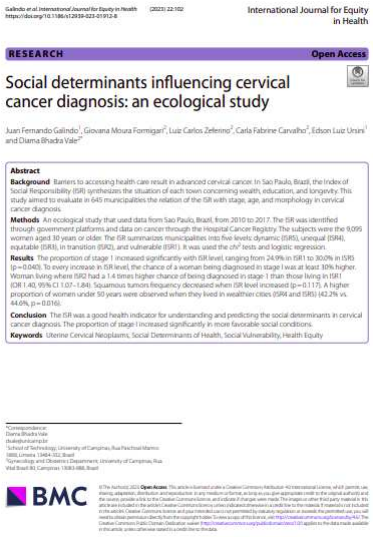
This work aims to use SDoH for cervical cancer to create an Agent-Based Model (ABM) of the evolution of cervical cancer among women to define strategies for the maximization of early-stage detection. The sources for the SDoH are two databases of the State of São Paulo, Brazil. A Reinforcement Learning algorithm (Q-Learning) is used to identify strategies.

2 MATERIALS AND METHODS

For the construction of the model, different stages are being performed. In the first stage, a Logit model was created fusing SDoH For ABM validation. In the second stage, the ABM was created. In the third stage, a Q-Learning algorithm will be implemented to maximize the rate of detection at an early stage.

In the first stage, two databases were used. The first dataset comes from the São Paulo Oncological Foundation (FOSP), containing 9502 records and six variables of cancer diagnosis and treatment data of hospitals in the state of Sao Paulo. The second dataset is published by the State of São Paulo Statistics portal (SEADE). It contains 645 records and 17 variables representing social, economic, and demographic data from the cities belonging to the state. The two datasets were combined. Then, data imputation was used to complete the missing variables (Jaramillo et al. 2021).

Figure C.4: Use of Social Determinants of Health in Agent-based Models Abstract



Galindo et al. International Journal for Equity in Health (2023) 22:102
https://doi.org/10.1186/s12939-023-01912-8

International Journal for Equity in Health

RESEARCH Open Access

Social determinants influencing cervical cancer diagnosis: an ecological study

Juan Fernando Galindo[1], Giovana Moura Formigari[2], Luiz Carlos Zeferino[2], Carla Fabrine Carvalho[2], Edson Luiz Ursini[1] and Diama Bhadra Vale[2*]

Abstract

Background Barriers to accessing health care result in advanced cervical cancer. In Sao Paulo, Brazil, the Index of Social Responsibility (ISR) synthesizes the situation of each town concerning wealth, education, and longevity. This study aimed to evaluate in 645 municipalities the relation of the ISR with stage, age, and morphology in cervical cancer diagnosis.

Methods An ecological study that used data from Sao Paulo, Brazil, from 2010 to 2017. The ISR was identified through government platforms and data on cancer through the Hospital Cancer Registry. The subjects were the 9,099 women aged 30 years or older. The ISR summarizes municipalities into five levels: dynamic (ISR5), unequal (ISR4), equitable (ISR3), in transition (ISR2), and vulnerable (ISR1). It was used the chi^2 tests and logistic regression.

Results The proportion of stage I increased significantly with ISR level, ranging from 24.9% in ISR1 to 30.0% in ISR5 ($p = 0.045$). To every increase in ISR level, the chance of a woman being diagnosed in stage I was at least 30% higher. Woman living where ISR5 had a 1.4 times higher chance of being diagnosed in stage I than those living in ISR1 (OR 1.40, 95% CI 1.07–1.84). Squamous tumors frequency decreased when ISR level increased ($p = 0.117$). A higher proportion of women under 50 years were observed when they lived in wealthier cities (ISR4 and ISR5) (42.2% vs. 44.6%, $p = 0.016$).

Conclusion The ISR was a good health indicator for understanding and predicting the social determinants in cervical cancer diagnosis. The proportion of stage I increased significantly in more favorable social conditions.

Keywords Uterine Cervical Neoplasms, Social Determinants of Health, Social Vulnerability, Health Equity

*Correspondence:
Diama Bhadra Vale
dvale@fcm.unicamp.br
[1]School of Technology, University of Campinas, Rua Paschoal Marmo, 1888, Limeira 13484-332, Brazil
[2]Gynecology and Obstetrics Department, University of Campinas, Rua Vital Brasil 80, Campinas 13083-888, Brazil

Figure C.5: Social determinants influencing cervical cancer diagnosis: an ecological study

C.6 AGENT-BASED MODEL FOR ANALYSIS OF CERVICAL CANCER DETECTION

Type: Extended Abstract. **Year:** 2023 **Authors:** Juan Fernando Galindo Jaramillo; Leonardo Grando; Jose Roberto Emiliano Leite; Diama Bhadra Vale; Edson Luiz Ursini. **Conference:** Winter Simulation Conference 2023. **Available at:** https://informs-sim.org/

Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

AGENT-BASED MODEL FOR ANALYSIS OF CERVICAL CANCER DETECTION

Juan F Galindo Jaramillo
Leonardo Grando
Jose Roberto Emiliano Leite
Diana Bhadra Vale
Edson Luiz Ursini

School of Technology - School of Medical Sciences University of Campinas
R. Paschoal Marmo, 1888
Limeira, SP 13484-332, BRAZIL

ABSTRACT

Using Agent-Based Models (ABM) for disease incidence may help decision-making processes. This work shows an ABM for cervical cancer detection. Our results show the relevance of social indicators.

1 INTRODUCTION

The analysis of disease incidence may help health organizations to improve patients' wellbeing. These analyses must consider Social Determinants of Health (SDoH), that is, the social and environmental conditions in the places people live (on Social Determinants of Health et al. 2008). In cancer care, spending is increasing, and some costs are related to SDoH (Yabroff et al. 2019). For cervical cancer, treatment costs are high, especially when detected at an advanced stage (Singh et al. 2020). The use of ABM eases modeling behaviors and their impact in the overall public or private system. For example, focus on modeling spread and countermeasures for COVID-19 may help decision-making (Hunt and Kain 2022).

This work shows the creation of an ABM for studying the early detection of cervical cancer considering SDoH in the State of São Paulo, Brazil. We previously found a relationship between SDoH and early detection (Galindo et al. 2022). Specifically, the Index of Social Responsibility (ISR) of the State of São Paulo, based on the Human Development Index. ISR goes from 1 (low wealth, schooling, and longevity) to 5 (high wealth, schooling, and longevity).

2 MODEL DESCRIPTION

We considered two types of agents: women and health facilities both located within a city. There are 64 cities with one facility per city. For validation, we performed a Logit regression model, with the cervical cancer detection stage as the dependent variable and women's age and schooling level, as well as cities' ISR, mean income rate, access to water, garbage, elementary school percentage, fecundity level, and beds per 1000 inhabitants as independent variables. We used data from the São Paulo Oncological Foundation (FOSP) and from the State of São Paulo Statistics (SEADE). We used these results to create an ABM to check the detection of cervical cancer at an early stage. To keep the model simple, from the variables present in the Logit model, we focused in this work on the ISR for cities.

Each time step corresponds to one year, when women try to be tested at their city's facility. Each facility decides to attend a number of women based on its capacity, related to the city's ISR. Also, each year, women may develop cancer. As the number of years without attention increases, the chances to have cancer at an advanced stage increase. For initial conditions, women's location is random, and cancer and

Figure C.6: AGENT-BASED MODEL FOR ANALYSIS OF CERVICAL CANCER DETECTION